Convergent Numerical Methods for Nonlinear Partial Differential Equations

Franziska Leonie Weber

Dissertation presented for the degree of Philosophiae Doctor

Department of Mathematics
University of Oslo

June 2015

Series of dissertations submitted to the
Faculty of Mathematics and Natural Sciences, University of Oslo
No. 1647

ISSN 1501-7710

Cover: Hanne Baadsgaard Utigard.
Printed in Norway: AIT Oslo AS.

Produced in co-operation with Akademika Publishing.
The thesis is produced by Akademika Publishing merely in connection with the thesis defence. Kindly direct all inquiries regarding the thesis to the copyright holder or the unit which grants the doctorate.

Acknowledgments

This thesis was written at the Department of Mathematics at University of Oslo and funded by NRF research project 214495 LIQCRY. The Norwegian Research Council is acknowledged for the generous financial support.

I wish to thank my advisors, Prof. Dr. Nils Henrik Risebro and Prof. Dr. Helge Holden for encouraging me to do a Ph.D., and for many interesting discussions and invaluable advice on scientific matters as well as survival in Norway.

I am also very grateful to my collaborators, Prof. Dr. Giuseppe M. Coclite, Prof. Dr. Trygve K. Karper, Prof. Dr. Siddhartha Mishra, Prof. Dr. Christoph Schwab and Prof. Dr. Konstantina Trivisa, for suggesting interesting problems, helping me tackling them, and thereby teaching me a lot. I greatly enjoyed working with you. Without you, this thesis would never have been finished!

In fall 2014, I had the opportunity to visit CSCAMM at University of Maryland, which was a great experience for me. I had the opportunity to interact with various gifted mathematicians in this inspiring research environment and learned a lot. I am deeply grateful to Prof. Dr. Eitan Tadmor for the invitation and to CSCAMM for the hospitality.

Many thanks go to my wonderful friends and to my collegues at Blindern for spending time with me on weekends or during coffee breaks, listening to my complaints about the research and motivating me to keep going when I was frustrated, skiing and running with me, and joining me for other non-mathematical activities. I am so glad to know you!

Last but not least, I wish to thank my family, in particular my parents, for all the support, patience, and understanding over the years.

Contents

Introduction

In the 17th century, Newton and Leibnitz invented calculus and so laid the fundamentals for the subject of partial differential equations (PDEs). Relating the state of a physical system to its neighboring states in space-time, these have become a powerful instrument to describe a variety of problems in engineering, physics, astronomy and biology using the formal language of mathematics. They provide a tool to obtain answers to questions such as "What is the acceleration needed to launch a rocket?", "Will this building resist the forces of a tornado or an earthquake?", "What will the weather be like tomorrow?", "How do we design an airplane to use as little fuel as possible?" by expressing the questions as the unknowns in equations relating derivatives and integrals of the unknown functions.

However, as John von Neumann said [**124**] "Truth (...) is much too complicated to allow anything but approximations"; nature is complex and various different processes influence each other, so that it is impossible to include all of them in a particular model. Simplifications need to be made and certain effects have to be neglected in the course of modeling. Often the resulting systems of PDEs are still far too complicated to be solved by hand and only for a relatively small class of PDEs can we find explicit formulae, expressing their solutions in terms of the data. Thus, further approximations are required, the PDE has to be *discretized*. This means that the differential equation, the *continuous* problem, is cast into a finite system of algebraic equations with a finite number of unknowns, the *discrete* problem, which then can be implemented and solved using a computer.

In this process of approximation, there are several sources of error:

(i) The models are based on empirical observations and experiments and are therefore subject to *measurement errors*. Inaccurate measurements of physical parameters can affect the models derived based on them.

(ii) When formulating a model, many physical effects have to be neglected in order to keep the model comprehensible. How does one decide on whether a certain effect is not relevant and can be disregarded? Or why is another one important enough to be included in the model? *Modeling errors* could be made.

(iii) By approximating the solution to the PDE numerically, we make a *discretization error*. Can we be sure the finite-dimensional output our numerical method provides is an accurate approximation to the function which is the actual solution of the PDE?

Numerical analysts are typically concerned with the third issue. Their task is to develop efficient and reliable numerical algorithms that can be implemented on a computer to

generate good approximations to the true solution. Then the algorithms can be used to test the hypotheses made in the modeling step.

To comply with this task, the discretization schemes have to be designed so that they satisfy certain mathematical properties which guarantee that the simulations are accurate and that the approximations are close to the actual solution of the PDE. In mathematical terms, we speak about stability, convergence of the method, and error estimates, which quantify how close the approximations are to the solution.

If the equations are nonlinear, which is the case in many applications, ensuring that the discretizations satisfy these properties is more involved. Solutions to such equations often exhibit complex structures. They develop shock waves, rapid oscillations, and blow-ups. For this reason, they are not differentiable in the classical sense, and weaker, more general notions of differentiability and solution need to be defined. This complicates the task of devising stable but still efficient numerical schemes. In addition, establishing convergence of the sequence of approximations to the solution and quantifying the error in the approximations becomes more elaborate due to the lack of differentiability of solutions. Fortunately, the explicit nonlinear structure often provides extra information that facilitates proving convergence to the solution. The goal is therefore to construct numerical methods which preserve a discrete version of these special structures.

In this thesis, we design and analyze such discretizations for various applications. Specifically, we construct fully-discrete finite difference schemes for two-phase flow in porous media; a mathematical model for tumor growth; and the wave map equation into the sphere. We prove that the approximations defined by the numerical schemes converge to the solutions of the respective PDEs, and hence know that we have control over the discretization error, we can make it arbitrarily small by increasing the computational effort. We test the performance of the schemes in numerical experiments and discuss their ability to capture physical properties of the phenomena they attempt to model. The applications are all related to wave and transport phenomena in fluids, but the nonlinearities manifest themselves in different ways: In Paper 1, the constraint that the solution map takes values in the sphere causes a quadratic nonlinearity in the gradient and in the numerical experiments we observe blow-ups of the gradient while the solution itself stays bounded thanks to the constraint. In Paper 2, the nonlinearity comes from the velocity in the transport equation and we have to deal with oscillatory, unbounded solutions. In Article 3, we have again a nonlinearity in the transport velocity and additionally a nonlinear source term. The solutions to that system of PDEs are bounded but develop sharp gradients.

The fourth paper of the thesis is more closely related to the first two error sources, the measurement and modeling errors: We represent the data and modeling uncertainty as random parameters in the equations and compute approximations to the resulting random hyperbolic conservation law using a combination of front tracking and multilevel Monte Carlo methods to deal with the randomness. Front tracking has been used successfully as a numerical method to approximate hyperbolic conservation laws, and error estimates are available [**73**]. Thus the discretization error can be controlled. Similarly, multilevel Monte Carlo methods [**55**] have been widely used for the simulation of stochastic PDEs.

Combining the two, we derive error estimates in terms of the moments of the solution to the random hyperbolic conservation law.

Since the applications as well as the challenges faced due to the nonlinearities in the equations differ from paper to paper, we describe in the following the results of each of them separately.

1. Summary: A New Angular Momentum Method for Computing Wave Maps into Spheres

This paper is a joint work with Trygve K. Karper and was published in SIAM Journal of Numerical Analysis, 2014 [**79**].

1.1. Background. *Wave maps* can be motivated in two ways: On the one hand, they can be seen as a generalization of the wave equation in Euclidean space to a wave equation taking values on a manifold. On the other hand, they can be considered as harmonic maps on Lorentzian manifolds [**118**]. Let us start by explaining the first approach. The wave equation

$$u_{tt} = \Delta u, \quad (t,x) \in \mathbb{R}^+ \times \Omega,$$

where $\Omega \subset \mathbb{R}^n$, is a model to describe the motion of an n-dimensional surface in an ambient Euclidean space. The one-dimensional linear wave equation was first formulated by d'Alembert in 1746 [**112**], when he was studying the vibration of strings like, for example, those used for musical instruments. As the name "wave equation" indicates, the linear wave equation and related equations describe the motion of waves occuring in physics, such as water waves, sound, electromagnetic and light waves. One can then think of the wave map equation as describing the motion of an n-dimensional object or wave front on a manifold. Alternatively, let us have a look at the Laplace equation

$$-\Delta u = 0, \quad x \in \Omega.$$

Its solutions, called *harmonic functions*, can be thought of as critical points of the Lagrangian

$$\mathcal{L}^e(u) = \frac{1}{2}\int_\Omega |\nabla_x u|^2 dx.$$

If we now replace $\mathbb{R}^n$ by a general Riemannian manifold M and consider functions $u : \Omega \to M$ taking values in M, then its derivatives are elements of the tangent space $T_u M$ of M and we can study critical points of the energy

$$\mathcal{L}^e_M(u) = \frac{1}{2}\int_\Omega |\nabla_x u|_g^2 \, dx, \tag{1.1}$$

where $|\nabla_x u|_g^2 = \langle \nabla_x u, \nabla_x u\rangle_g$ and $\langle \cdot, \cdot \rangle_g$ denotes the inner product induced by the metric g on the manifold M. Solutions to the corresponding Euler-Lagrange equation are called *harmonic maps* and have been studied extensively [**65, 66**]. Instead of minimizing the energy (1.1), we can form the action functional

$$\mathcal{L}^h_M(u) = \frac{1}{2}\int_0^\infty \int_\Omega \left(-|\partial_t u|_g^2 + |\nabla_x u|_g^2\right) dxdt,$$

and compute its Euler-Lagrange equation,

$$\nabla_t^M \partial_t u = \sum_{i=1}^{n} \nabla_{x_i}^M \partial_{x_i} u, \quad (t,x) \in \mathbb{R}^+ \times \Omega, \tag{1.2}$$

where ∇_{t,x_i}^M are the covariant derivatives. Solutions to this equation are called *wave maps*. The two terms in the Lagrangian $\mathcal{L}_M^h$ can be interpreted as the difference between potential and kinetic energy. On a formal level, the two types of energy stay in balance since the total energy

$$\mathcal{E}(t) := \frac{1}{2} \int_\Omega |\partial_t u(t,\cdot)|_g^2 + |\nabla_x u(t,\cdot)|_g^2 \, dx,$$

stays constant in time, as it is the case for the linear wave equation on Euclidean space. Moreover, as their linear version, wave maps have a finite speed of propagation property. However, even though equation (1.2) looks harmless at first sight, its solution can develop singularities due to the nonlinearity which is hidden in the covariant derivatives. If the target manifold M is the sphere embedded in Euclidean space, we can rewrite equation (1.2) as

$$u_{tt} - \Delta u = \left(|\nabla u|^2 - |u_t|^2 \right) u, \quad |u| = 1, \tag{1.3}$$

which is the equation which we study in our paper in the case of the sphere embedded in $\mathbb{R}^3$. In this formulation, the nonlinearity in the first order derivatives is obvious. Wave maps have been studied extensively by Struwe and Shatah [**111**], Tataru [**119, 118**], Tao [**115, 116**], Krieger and Schlag [**82**], and others. However, there are still major gaps in the understanding of this type of equation, for example uniqueness of weak solutions is still an issue [**114, 127**].

The wave map equation can be seen as a simplified model for Einstein equations of general relativity and is related to the Yang-Mills equations, [**117**].

1.2. Summary of results. In our paper, we develop a finite difference method based on a reformulation of (1.3). Specifically, we introduce the *angular momentum* $w := u_t \times u$ and rewrite (1.3) in terms of w as a system of two equations:

$$\begin{aligned} u_t &= u \times w, \\ w_t &= \Delta u \times u. \end{aligned} \tag{1.4}$$

In this formulation, the constraint $|u| = 1$ is inherent and hence there is no need for a Lagrange multiplier. This can be seen by taking the inner product of the first equation with u. For the time discretization, we use the midpoint rule,

$$\begin{aligned} D_t u^m &= u^{m+1/2} \times w^{m+1/2}, \\ D_t w^m &= \Delta u^{m+1/2} \times u^{m+1/2}. \end{aligned} \tag{1.5}$$

This time integration satisfies $|u^{m+1}| = |u^m|$ and at the same time conserves the energy, that is,

$$\int_\Omega |w^{m+1}|^2 + |\nabla u^{m+1}|^2 \, dx = \int_\Omega |w^m|^2 + |\nabla u^m|^2 \, dx.$$

To discretize in space, we use a standard discretization of the Laplace operator on a rectangular grid. Since (1.5) is nonlinear and implicit, we use a fixed point solver to approximate its solution. We show that the fixed point solver converges under a linear CFL-condition on the time step $\Delta t \leq Ch$ with respect to the spatial grid width h and finally prove that approximations computed with the method (1.5) converge to weak solutions of the wave map equation (1.3).

1.2.1. *Related equations and methods.* Writing the wave map equation in the form (1.4), one can see that it is related to the Landau-Lifshitz-Gilbert (LLG) equation

$$M_t = \alpha M \times H_{\text{eff}} + \beta M \times (M \times H_{\text{eff}}),$$

which is used in micromagnetics to model the effects of a magnetic field on a ferromagnetic material. The LLG equation predicts the rotation of the magnetization M in response to torques in a ferromagnet. The magnitude of M is equal to the saturation magnetization M_s at each point. The *effective field* H_{eff} is given as $H_{\text{eff}} = -\frac{\delta \mathcal{E}}{\delta M}$, where $\mathcal{E}$ is the free energy,

$$H_{\text{eff}} = \kappa \Delta M + F(M).$$

Here κ is constant, and $F(M)$ are lower order terms [**18**]. Numerical methods for this model were developed by Krishnaprasad and Tan [**83**], and Bartels and Prohl [**11**]. Other related equations occur in rigid body systems, as for example the one investigated by Austin, Krishnaprasad and Wang [**4**].

1.3. Future research directions and open questions. Our original motivation to consider wave map equations stems from the study of liquid crystal flows. These are materials which exhibit intermediate states between the liquid and the solid phase. They typically consist of elongated molecules that tend to align along a preferred axis. In the so-called Oseen-Frank model [**50**], this main orientation of the molecules is described by the *director field* $u(x) : \Omega \to \mathbb{S}^2$ and their tendency to align along the same axis is characterized by the Oseen-Frank elastic energy [**113**],

$$W_{OF} = \frac{1}{2}\left(k_1(\operatorname{div} u)^2 + k_2(u \cdot (\operatorname{curl} u))^2 + k_3(u \times (\operatorname{curl} u))^2\right),$$

where k_1, k_2, and k_3 are material constants. If one assumes that the flow velocity is zero and that inertial effects dominate viscosity, the dynamics of the liquid crystal director u can be described by the Euler-Lagrange equations corresponding to the least action principle [**110**],

$$\mathcal{L}(u) = \int_0^\infty \int_\Omega \frac{1}{2}|u_t|^2 - W_{OF}(u, \nabla u)\, dx\, dt.$$

The wave map equation corresponds to the special case when $k_1 = k_2 = k_3 = 1$, the *one-constant approximation*. The finite difference method developed in our paper could easily be extended to a stable energy and constraint preserving method for more general coefficients k_1, k_2, and k_3, however, proving convergence to a weak solution of the corresponding Euler-Lagrange equation could turn out to be very challenging. Another interesting project would be to extend the angular momentum method to a system of liquid crystal director

coupled to an external electric field or to the flow of the surrounding fluid and prove convergence of the resulting method to a weak solution of the system; and finally to extend it to the full Ericksen-Leslie equations.

Alternatively, one could try to extend the method to a liquid crystal model with variable degree of orientation, like the one investigated by Calderer *et al.* in [**22**].

2. Summary: Analysis and Numerical Approximation of Brinkman Regularization of Two-Phase Flows in Porous Media

This paper is a joint work with Giuseppe M. Coclite, Siddhartha Mishra and Nils Henrik Risebro and was published in Computational Geosciences, 2014 [**31**].

2.1. Background. In this paper, we investigate a model for two-phase flow in porous media. The modeling of such subsurface flow phenomena is important for many applications, in particular for the simulation of petroleum reservoirs. Roughly speaking, a petroleum reservoir is a permeable porous rock containing hydrocarbons. These sediments settled circa 10 million years ago, got buried and compressed and during these millions of years evolved into the form they have nowadays. Oil and petroleum gas constitute important sources of energy for our society, and for this reason we are interested in extracting the oil from beneath the surface.

Initially, the reservoir may be under enough pressure to push the hydrocarbons to the surface. But as the pressure declines, water or gas has to be injected to maintain the pressure and push more oil to the surface, this is the so-called *secondary recovery* and constitutes a typical example of two- or multi-phase flow in porous media. Since one would like to predict relative flow rates and directions in which the fluids flow under the surface in order to maximize the oil production, the simulation of reservoirs using computer models becomes crucial.

2.2. Summary of results. We consider a simple model of two-phase flow which consists of a transport equation coupled with the Brinkman regularization of Darcy's law:

$$\begin{aligned} \partial_t s + \operatorname{div}_{\mathbf{x}} (\mathbf{v}_w) &= 0, \\ -\mu \Delta_{\mathbf{x}} \mathbf{v}_w + \mathbf{v}_w &= -f(s)\lambda_T(s)\nabla_{\mathbf{x}} p, \\ -\operatorname{div}_{\mathbf{x}} (\lambda_T(s)\nabla_{\mathbf{x}} p) &= 0. \end{aligned} \tag{2.1}$$

We define a suitable notion of weak solution and prove existence of it via a vanishing viscosity approximation and via a numerical approximation by a finite difference scheme. The scheme's robustness for a fixed parameter $\mu > 0$ in the Brinkman regularization is demonstrated in a couple of experiments and then the limit as $\mu \to 0$ is investigated numerically using the same scheme. Formally setting $\mu = 0$, we would recover Darcy's law in (2.1). However, our stability estimates on the saturation and the velocity blow up as $\mu \to 0$ and at the same time, the numerical approximation becomes very oscillatory and the saturation attains nonphysical values. This indicates that the solutions of the Brinkman regularization (2.1) might not converge to solutions of Darcy's law as $\mu \to 0$. A complementary analysis in one space dimension further consolidates this conjecture.

2.3. Future research directions and open questions. In the one-dimensional case, we were able to prove uniqueness of (classical) solutions, in the multi-dimensional case, this is still an open issue. Proving this could be a preliminary step to a stability result with respect to the parameter μ. Another interesting problem might be to set the viscosity parameter ε in the vanishing viscosity approximation and the parameter μ in the Brinkman law in relation, try to pass to the limit in both of them at the same time, and investigate traveling wave solutions depending on the two parameters. This is harder than finding traveling waves for system (5.1) (in the article, where $\varepsilon = 0$), as the resulting system of ODEs will consist of three equations. In our paper, we made the assumption that the capillary pressure is zero. The inclusion of it into the model and an analysis of the resulting effects could be another topic for further investigation.

3. Summary: A Convergent Explicit Finite Difference Scheme for a Mechanical Model for Tumor Growth

This paper is a joint work with Konstantina Trivisa.

3.1. Background. In recent years, there has been an increasing interest in the mathematical modeling and numerical simulation of tumor growth to complement experimental and clinical studies and thereby improve the understanding of cancer development. Nowadays, a large number of models describing cell multiplication in a tissue is available. They can essentially be divided into two subgroups: Individual cell (agent)-based models, see for example [**42**], and continuum models which can be used for the description of large scale solid tumors. In the individual cell-based models the basic unit is the cell, whereas the continuum models either describe the dynamics of the cell population density [**21**] or the geometric motion of the tumor through a free boundary value problem, see [**33, 35, 34, 52**]. For a comparison of the two theoretical approaches, we refer to [**20**].

A first class of models for the growth of solid tumors was developed by Greenspan [**62, 61**]. They are based on the assumption that in a first stage, when the tumor is still small, the growth rate of the population is exponential since in this stage every cell receives enough nutrients (oxygen and glucose) due to diffusion (*avascular growth*). As the number of cells increases, the nutrient concentration decreases in the center of the tumor and when it falls below the critical level to sustain cell life, a necrotic core develops. This happens when the size of the tumor reaches approximately 1mm. The tumor then secretes diffusible substances, so-called tumor angiogenesis factors (TAF) into the surrounding tissue which trigger the development of neovasculatures to supply the tumor with enough nutrients (*vascular growth phase*) [**24**].

This motivated the development of a new generation of models where the growth is limited by the competition for space [**19**]. Such descriptions are based on mechanical concepts, considering the tissue as a multi-phase fluid. The phases can for example be water, healthy and tumor cells, extra-cellular matrix etc. Using such a description, the ability of the tumor to expand into a host tissue is primarily driven by the cell division rate which depends on the local cell density and the mechanical pressure in the tumor. As soon as the pressure reaches a critical level, termed *homeostatic pressure*, cell mulitplication is

prevented due to contact inhibition. In mathematical terms, this can be described using a transport equation with a source term

$$\partial_t n - \operatorname{div}(n\boldsymbol{u}) = n\boldsymbol{G}(p), \tag{3.1}$$

where n represents the number density of tumor cells, $\boldsymbol{u}$ the velocity field and p the pressure of the tumor. The term $n\boldsymbol{G}(p)$ on the right hand side expresses the growth of the cell culture in relation to the pressure. Following [**20, 108**], we assume that $\boldsymbol{G}$ is of the form

$$\boldsymbol{G} \in C^1(\mathbb{R}), \quad \boldsymbol{G}'(\cdot) \leq -\beta < 0, \quad \boldsymbol{G}(P_M) = 0 \quad \text{for some } P_M > 0. \tag{3.2}$$

P_M is the homeostatic pressure, the critical threshold at which the cell division is stopped by contact inhibition. It is related to the compression a cell can experience [**20**]. Here, we assume that the pressure is an increasing function of the cell density n, specifically, we will assume

$$p(n) = an^\gamma, \tag{3.3}$$

where $\gamma \geq 2$. Due to proliferation and removal of cells, there is a continuous motion within the tumor represented by a velocity field $\boldsymbol{u}$. We make the assumption that this velocity is given as a potential $\boldsymbol{u} = \nabla W$, where W is the solution to Brinkman's equation

$$p = -\mu\Delta W + W \tag{3.4}$$

for a viscosity coefficient $\mu > 0$. If viscosity is neglected, that is $\mu = 0$, we recover Darcy's law and (3.1) can be rewritten as the porous media equation with a source term. A detailed analysis of that equation can be found in the monograph [**123**]. In that case, only the friction of the tumor cells with the surrounding extra-cellular matrix is considered. The viscosity term $\mu\Delta W$ is therefore a way to take the friction between the cells themselves into account, considered as a Newtonian fluid [**107**].

3.2. Summary of results. In our paper we are concerned with the numerical approximation of the system (3.1)-(3.4) on a bounded domain Ω. We define a suitable notion of weak solutions to the system of equations and show existence of such solutions via a vanishing viscosity approximation and via the approximation by a finite difference scheme. In order to establish compactness of the approximating sequences, we prove some a priori estimates that give us uniform boundedness of the cell density n and $H^2(\Omega)$-regularity for the potential W. In order to prove strong convergence of the cell density n, we use a monotonicity property of the source term $n\boldsymbol{G}(p)$ which we combine with a compensated compactness argument for the pressure. We present a few numerical examples to demonstrate the performance of the difference scheme.

3.2.1. *Related work.* Related work on the mathematical analysis of mechanical models of Hele-Shaw-type have been presented by Perthame *et al.* [**104, 105, 106, 107**]. The analysis in [**106**] establishes the existence of traveling wave solutions of the Hele-Shaw model of tumor growth with nutrient and presents numerical observations in two space dimensions. In [**105, 107**], the limit $\gamma \to \infty$ in the pressure law (3.3) is investigated, however, no rigorous proof of existence of solutions to (3.1)-(3.4) is given. The present

article is according to our knowledge the first article presenting rigorous results on the global existence of weak solutions to (3.1)-(3.4).

A different approach yielding results on the global existence of weak solutions to a nonlinear model for tumor growth in a general moving domain $\Omega_t \subset \mathbb{R}^3$ without any symmetry assumption and for finite large initial data is presented in [**41**].

Relevant results on nonlinear models for tumor growth governed by the Darcy's law for the evolution of the velocity field are presented by Zhao [**128**] based on the framework introduced by Friedman *et al.* [**51, 26**]. The analysis in [**51, 26**] yields existence and uniqueness of solution to a related model in the radial symmetric case for a small time interval $[0, T]$. The analysis in [**128**] treats a parabolic-hyperbolic free boundary problem and provides a unique global solution in the radially symmetric case.

3.3. Future research directions and open questions. As a next step, we would like to show uniqueness of solutions to (3.1)-(3.4). Furthermore, we are planning to extend the difference scheme discussed in our paper to more realistic models which include the effects of nutrients, e.g. [**105**], and drugs; and to more general boundary conditions. Another interesting project would be the design of a numerical method for a model with moving domain, as for example the one in [**41**].

In our article there was a discrepancy concerning the regularity of the continuous potentials W_ε of the vanishing viscosity approximation and the discrete potentials W_h obtained via the numerical approximation: Whereas we could show that $W_\varepsilon \in L^\infty([0,T], W^{2,p}(\Omega))$, for any $1 \leq p < \infty$ using Calderón-Zygmund inequality, we were only able to show that $W_h, \nabla_h W_h, \nabla_h^2 W_h \in L^\infty([0,T]; L^2(\Omega))$ which corresponds to $H^2(\Omega)$-regularity in the continuous case. Improving this regularity result (maybe via a discrete version of the Calderón-Zygmund inequality) would be nice as it would improve the time step restriction for the numerical scheme to a linear condition.

3.4. Comparison of the models in Papers 2 and 3. At first sight, the models analyzed in Articles 2 and 3 look very similar: Both consist of a transport equation where the velocity is given by Brinkman's law. Nevertheless, the estimates which we obtain for the saturation s and the tumor cell density n differ: In [**31**], we show that the saturation $s \in H^1([0,T]; H^1(\Omega))$, but with estimates depending on the parameter μ in Brinkman's law, whereas for the tumor cell density n we can only show $L^\infty((0,T)\times\Omega)$ bounds (however, uniform in μ) and we need to use compensated compactness arguments to conclude strong convergence of the approximating sequences. The oscillations observed in the numerical experiments of Paper 2 indicate that the saturation is not uniformly bounded in L^∞, while on the other hand the sharp fronts in the experiments in Paper 3 indicate that the cell density n might not have higher order regularity. The difference in behavior could be caused by the different pressure laws: Whereas in (2.1), the pressure is given as a solution of an elliptic equation, the pressure law in the model (3.1)-(3.4) is given as a power law of the density, (3.3). In addition, in model (2.1), the divergence term in the transport equation for the saturation s is the solution of Brinkman's equation, while in (3.1) the divergence term is the product of the cell density n with the velocity $\boldsymbol{u}$ given as the solution of Brinkman's equation (which helps us in obtaining the uniform L^∞-bound).

4. Summary: Mulitlevel Monte Carlo Front Tracking for Random Scalar Conservation Laws

This paper is a joint work with Nils Henrik Risebro and Christoph Schwab and was published in BIT Numerical Mathematics, 2015 [**109**].

4.1. Motivation. Many complex physical problems can be modeled by first order hyperbolic systems of conservation or balance laws. These can be written in the generic form

$$\begin{aligned} \mathbf{U}_t + \sum_{j=1}^{n} \frac{\partial}{\partial x_j}(\mathbf{F}_j(\mathbf{U})) = 0 \quad x = (x_1, \dots, x_n) \in \mathbb{R}^n,\ t > 0, \\ \mathbf{U}(x, 0) = \mathbf{U}_0(x), \quad x \in \mathbb{R}^n, \end{aligned} \tag{4.1}$$

where $\mathbf{U} : \mathbb{R}^n \to \mathbb{R}^m$ is the vector of unknowns and $\mathbf{F}_j : \mathbb{R}^m \to \mathbb{R}^m$ is the flux vector for the j-th direction with m being a positive integer.

This type of partial differential equation is ubiquitous; the shallow water equations of hydrology, the Euler equations for inviscid, compressible flow and the magnetohydrodynamic (MHD) equations of plasma physics, see for example [**37, 58**], fall into this category, to mention a few. In our article we focus on the scalar case $m = 1$ in (4.1), which is termed *scalar conservation law* (SCL).

Even though solutions to hyperbolic conservation laws develop discontinuities (*shock waves* and *contact discontinuities*) in finite time, yet when the initial data is smooth, and are therefore not differentiable in the classical sense, the theory of scalar conservation laws is well-studied [**37, 58**]. A weak notion of solutions, in which (4.1) is interpreted in the distributional sense, has been introduced, and the framework of *entropy solutions*, in which equation (4.1) is augmented with additional *entropy conditions*, has been developed to establish uniqueness of solutions in the space of integrable functions $L^1(\mathbb{R}^n)$ [**85**].

Furthermore, numerical methods to approximate entropy solutions of systems of conservation laws such as (4.1) have undergone extensive developement and nowadays, a large selection of efficient and stable numerical schemes are available, see for example [**48, 58, 59, 91**] and the references therein.

However, this classical paradigm for devising efficient numerical methods assumes that the input data, here the initial condition $\mathbf{U}_0$ and the flux function $\mathbf{F}_j$, are known exactly. In many situations of practical interest, this is in fact not the case. These data are obtained conducting empirical experiments and measurements of physical parameters and are thus subject to modeling uncertainty and measurement errors. The real question is then: How accurate are our results?

Thus we would like to quantify, how much errors and uncertainty in the input data will affect the solution of (4.1). In order to answer this question in a proper mathematical framework, we allow for *random data*, specifically, we replace the deterministic initial data $\mathbf{U}_0$ and the flux function $\mathbf{F}_j$ by random fields taking values in a function space. We are now dealing with a *random conservation law* where the unknown $\mathbf{U}$ is a random field. The numerical approximation of the solution $\mathbf{U}$ then occurs on two levels: As before, we need

to discretize the physical domain, but now, we need to approximate in addition in the stochastic space.

4.2. Summary of results. In our paper we combine a mulitlevel Monte Carlo (MLMC) method for the approximation in the stochastic space with a front tracking method for the physical space. Our work is based on [**97**] where the problem of random initial data was considered and the existence and uniqueness of a random entropy solution was shown, and a convergence analysis for multilevel Monte Carlo finite volume discretizations was given. We generalize this wellposedness result to allow for random flux functions. In order to achieve this, we use the concepts of *strongly measurable functions* and *Lebesgue-Bochner spaces* [**122**] to deal with the nonseparability of the spaces of Lipschitz functions and functions of bounded variation which are the natural spaces for the flux function and the initial condition to obtain wellposedness. The use of the front tracking method [**73**] for the approximation in the physical space improves in one space dimension the convergence rate of the MLMC front tracking method versus discretization parameter and versus work, in comparison to the MLMC finite volume method from [**97**]. In several space dimension, the asymptotic order of the convergence rate of the method is the same as the one for the MLMC finite volume method, but in contrast to that one, there is no CFL-restriction on the time step in the front tracking algorithm.

4.3. Future research directions. It might be interesting to extend and apply the method to conservation laws with spatially dependent flux functions, where for example the flux function contains a coefficient which is dependent on the material properties. This coefficient could be modeled as a random field and stability of the equation with respect to its perturbations could be investigated.

5. Perspectives

We have examined numerical methods for four different nonlinear problems which required a broad spectrum of techniques to prove convergence of the methods to the solutions of the respective PDEs.

In the first paper, it was essential to ensure that the approximations preserve the unit length constraint and a discrete version of the energy to establish convergence of the approximating sequences. The resulting method turned out to be implicit whereas the schemes developed in the other papers are explicit. An iterative solver was needed in practice to solve the discretized system of equations, but the method remained efficient thanks to a linear time step restriction.

In the second article, a Brinkman modification of Darcy's law enabled us to prove more regularity on the solution and by mimicking the corresponding estimates in the discrete setting, we were able to prove convergence of the approximations to the solution of the system of equations.

In the third paper, the velocity in the transport equation is again determined by Brinkman's law. In contrast to the solutions of the equations in Article 2, the solutions to the here considered system of equations are uniformly bounded but develop sharp fronts.

To prove convergence of the approximations, we used a compensated compactness argument for the convergence of the pressure combined with a monotonicity argument for the source term in the transport equation. A discretized version of this proof yielded convergence of the finite difference scheme.

Whereas the first three papers of the thesis are mainly related to controlling the discretization error, the focus was on the measurement and modeling errors in the fourth paper. Front tracking and mulitlevel Monte Carlo methods, which we used for the discretization of the physical and stochastic space respectively, are both well-known and precise error estimates with respect to the discretization parameters are available. We were therefore in a position to investigate how measurement and modeling errors affect the solutions.

We hope that the developed schemes and techniques can be applied to other related problems in the future. The problems discussed in the first three papers could for example be put in the stochastic setting of the fourth paper to gain more insight into the measurement and modeling errors made when deriving the equations, by combining the constructed numerical schemes with the multilevel Monte Carlo method from the fourth paper or with another stochastic method. This way, we might be able to improve the models to make better predictions about the phenomena they seek to describe.

PAPER 1

A New Angular Momentum Method for Computing Wave Maps into Spheres

Joint work with Trygve K. Karper

Abstract. In this paper, we present and analyze a new finite difference method for computing three dimensional wave maps into spheres. By introducing the angular momentum as an auxiliary variable, we recast the governing equation as a first order system. For this new system, we propose a discretization that conserves both the *energy* and the *length constraint*. The new method is also fast requiring only $N \log N$ operations at each time step. Our main result is that the method converges to a weak solution as discretization parameters go to zero. The paper is concluded by numerical experiments demonstrating convergence of the method and its ability to predict finite time blow-up.

1. Introduction

The purpose of this paper is to develop a new numerical method for computing wave maps. By wave maps, we here mean vectors $d = [d_1, d_2, d_3]^T$ satisfying the following constrained wave equation:

$$d_{tt} - \Delta d = \gamma d, \quad |d| = 1, \quad \text{in } (0,\infty) \times \Omega. \tag{1.1}$$

Here, $\Omega \subset \mathbb{R}^n$, $n = 2, 3$, is either assumed to be the unit box $\Omega = [0,1]^n$ or it is assumed that $\Omega = \mathbb{T}^n$, where $\mathbb{T}^n$ is the n-dimensional torus. In the first case, (1.1) is augmented with homogenous Neumann boundary conditions.

The γ appearing in (1.1) is a Lagrange multiplier enforcing the constraint $|d| = 1$. In particular, by dotting (1.1) with d and using that $|d| = 1$, one finds that

$$\gamma = |\nabla d|^2 - |d_t|^2.$$

Thus, (1.1) is in this sense highly nonlinear which in turn obscures the task of developing conservative numerical methods. Moreover, in three spatial dimensions, it is known that solutions of (1.1) may blow-up [**111**]. Specifically, there is initial data for which the gradient ∇d develops singularities in finite time. Thus, solutions of the wave map equation are not smooth. We will return to the issue of blow-up in the numerical section (Section 5).

The literature on numerical methods for (1.1) seems to be confined to a handful of results. In the papers [**7, 9, 8**], the authors develops convergent splitting and relaxation methods. With these methods, (1.1) is either solved iteratively, using one evolution step and one projection step onto the sphere, or the constraint $|d| = 1$ is relaxed altogether. In

the paper [**10**], the wave map equation (1.1) is computed using an approximate Lagrange multiplier γ_h. The approximate γ_h is then designed such that the constraint $|d| = 1$ holds. This leads to a γ_h which depend nonlinearly and implicitly on the unknown d_h.

The method we will develop in this paper differs significantly from the previous methods. It is more related to the constraint preserving methods [**12, 11**] for computing *heat maps* into spheres. The key observation allowing us to deduce an energy and constraint preserving method is a new formulation of (1.1). Specifically, by introducing the *angular momentum*:

$$w = d_t \times d,$$

the wave map equation (1.1) can be recast in the form

$$d_t = d \times w, \tag{1.2}$$

$$w_t = \Delta d \times d. \tag{1.3}$$

In this formulation, the constraint $|d| = 1$ is inherent and hence there is no need for the Lagrange multiplier γ. Constraint preserving time integration for this system is easily derived. Here, we will use the first order integration:

$$\begin{aligned} \frac{d^{m+1} - d^m}{\Delta t} &= d^{m+1/2} \times w^{m+1/2}, \\ \frac{w^{m+1} - w^m}{\Delta t} &= \Delta d^{m+1/2} \times d^{m+1/2}, \end{aligned} \tag{1.4}$$

where $d^{m+1/2} = \frac{1}{2}(d^m + d^{m+1})$ and similarly $w^{m+1/2} = \frac{1}{2}(w^m + w^{m+1})$. This integration method satisfies $|d^{m+1}| = |d^m|$. Moreover, by dotting the second equation with $w^{m+1/2}$ and adding the first equation dotted with $\Delta d^{m+1/2}$, one obtains

$$\int_\Omega |w^{m+1}|^2 + |\nabla d^{m+1}|^2 \, dx = \int_\Omega |w^m|^2 + |\nabla d^m|^2 \, dx,$$

and hence the method also conserves the energy. To discretize (1.4) in space, we will use a standard central difference approximation of the Laplace operator on a regular grid.

The only potential downside of using the discretization (1.4) is that it is nonlinear and implicit and hence requires implementing a fixed point solver. Moreover, this fixed point solver should be such that at least the length constraint is conserved at every iteration. In Section 4, we will give the details on how such a solver may be constructed and prove that a fixed point may be computed (up to any tolerance in energy norm) using only $N \log N$ operations, where N is the number of degrees of freedom of d. In practice, finding a solution with tolerance N^{-2} requires only around $5 - 10$ iterations depending on the regularity of the underlying solution, but not on N. Note that there is not much point in decreasing the tolerance beyond N^{-2} as the discretization error of (1.4) will then dominate the error.

Our main theoretical result in this paper is that the new angular momentum method converges to a weak solution as discretization parameters go to zero. The proof of this fact will follow directly using energy arguments together with the observation that

$$d \times \Delta d = -\operatorname{div}(\nabla d \times d).$$

The remaining parts of this paper are structured as follows: In the upcoming section, we will properly define the new method and prove some basic properties. Then, in Section 3, we will prove that the method converges to a weak solution as discretization parameters go to zero. In Section 4, we will provide a way to compute the needed fixed point through an iterative procedure and prove that a fixed point may be obtained using only $N \log N$ operations. In Section 5, the paper is concluded by a series of numerical experiments illuminating some of the properties of the new method.

2. The angular momentum method

Given a number of degrees of freedom N, we set $M = N^{\frac{1}{n}}$, where $n = 2$, 3 is the spatial dimension, and assume that M is an integer. In the following, we will use $n = 3$, the modifications needed for the two-dimensional case are straight forward. Next, we let $h = 1/M$ and set the time step $\Delta t = \kappa h$, where κ is some constant. The domain Ω is then discretized in terms of the N points

$$x_{i,j,k} = (ih, jh, kh), \quad i, j, k = 0, \dots, M.$$

To simplify notation, we introduce the multiindex $\underline{i} \in \mathcal{I}_N := \{0, \dots, M\}^n$ such that we can write

$$x_{\underline{i}} = x_{i,j,k}.$$

We will approximate d at these points. Specifically,

$$d^m_{\underline{i}} \approx d(m\Delta t, x_{\underline{i}}).$$

Next, let $\mathbf{e}_1 := (1, 0, 0)$, $\mathbf{e}_2 := (0, 1, 0)$, and $\mathbf{e}_3 := (0, 0, 1)$. Using these vectors, we then define the forward and backward difference operators

$$D^+_j d_{\underline{i}} = \frac{d_{\underline{i}+\mathbf{e}_j} - d_{\underline{i}}}{h}, \qquad D^-_j d_{\underline{i}} = D^+_j d_{\underline{i}-\mathbf{e}_j},$$

respectively, for $j = 1, 2, 3$, and $\underline{i} \in \mathcal{I}_N$. The standard central Laplace discretization is then defined as

$$\Delta_h d_{\underline{i}} = \sum_{j=1}^{3} D^+_j D^-_j d_{\underline{i}}.$$

If we introduce the backward gradient $\nabla_h = [D^-_1, D^-_2, D^-_3]^T$ and forward divergence $\operatorname{div}_h v = D^+_1 v^{(1)} + D^+_2 v^{(2)} + D^+_3 v^{(3)}$, we have the identity

$$\operatorname{div}_h \nabla_h = \Delta_h,$$

which will be convenient in the upcoming analysis.

For the time discretization, we will use the notation

$$d^{m+1/2} := \frac{d^m + d^{m+1}}{2}, \quad D_t d^m = \frac{d^{m+1} - d^m}{\Delta t}.$$

To approximate the initial conditions, we shall use the operator

$$\Pi[f]_{\underline{i}} = \frac{1}{(h)^3} \int_{(i-0.5)h}^{(i+0.5)h} \int_{(j-0.5)h}^{(j+0.5)h} \int_{(k-0.5)h}^{(k+0.5)h} f(y)\, dy.$$

We are now ready to state the new method.

DEFINITION 2.1. Given initial data $d^0 \in H^1(\Omega)$, $d_t^0 \in L^2(\Omega)$, let

$$(d_{\underline{i}}^0,\ w_{\underline{i}}^0) = \left(\Pi[d^0]_{\underline{i}}, \Pi[d_t^0]_{\underline{i}} \times d_{\underline{i}}^0\right), \quad \forall \underline{i}.$$

Determine sequentially,

$$d_{\underline{i}}^m, w_{\underline{i}}^m, \qquad \forall \underline{i} \in \mathcal{I}_N, \quad m = 1, \ldots,$$

by solving the nonlinear system

$$D_t d_{\underline{i}}^m = d_{\underline{i}}^{m+1/2} \times w_{\underline{i}}^{m+1/2}, \tag{2.1}$$

$$D_t w_{\underline{i}}^m = \Delta_h d_{\underline{i}}^{m+1/2} \times d_{\underline{i}}^{m+1/2}. \tag{2.2}$$

We will now prove some fundamental properties of the new method. To this end, it will be convenient to extend the numerical solution to all of Ω. For this purpose, we shall use the piecewise constant extension:

$$\begin{aligned} d_h^m(x) &= d_{\underline{i}}^m, \quad x \in E_{\underline{i}}, \\ w_h^m(x) &= w_{\underline{i}}^m, \quad x \in E_{\underline{i}}, \end{aligned} \tag{2.3}$$

where $E_{\underline{i}} = [(i-1/2)h, (i+1/2)h) \times [(j-1/2)h, (j+1/2)h) \times [(k-1/2)h, (k+1/2)h)$, $\underline{i} = (i,j,k) \in \mathcal{I}_N$. Observe that the numerical method can then be written

$$D_t d_h^m = d_h^{m+1/2} \times w_h^{m+1/2}, \tag{2.4}$$

$$D_t w_h^m = \Delta_h d_h^{m+1/2} \times d_h^{m+1/2}, \tag{2.5}$$

where Δ_h is derived in the obvious way.

LEMMA 2.2. *There exists a unique numerical solution to the method posed in Definition 2.1. Moreover, the length is preserved*

$$|d_{\underline{i}}^m| = |d_{\underline{i}}^0| = 1, \quad \forall \underline{i}, \quad m = 0, \ldots, \tag{2.6}$$

and the energy is preserved

$$E_m = E_0, \quad m = 1, \ldots, \tag{2.7}$$

where the energy is defined as

$$E_m = \frac{1}{2} \int_\Omega |\nabla_h d_h^m|^2 + |w_h^m|^2 \, dx. \tag{2.8}$$

PROOF. The existence of a unique solution will be proved through a constructive iteration in Section 4. The proof can be found in Corollary 4.9.

That the length is conserved, (2.6), follows immediately from (2.1). Indeed, dotting with $d_h^{m+1/2}$ yields

$$D_t d_h^m \cdot d_h^{m+1/2} = 0.$$

Finally, to prove (2.7), we calculate

$$\begin{aligned} D_t E_m &= \int_\Omega w^{m+1/2} \cdot D_t w^m - \Delta_h d^{m+1/2} \cdot D_t d^m \, dx \\ &= \int_\Omega \left(\Delta_h d^{m+1/2} \times d^{m+1/2}\right) \cdot w^{m+1/2} \, dx \\ &\quad - \int_\Omega \left(d^{m+1/2} \times w^{m+1/2}\right) \cdot \Delta_h d^{m+1/2} \, dx = 0. \end{aligned}$$

This concludes the proof. □

3. The method converges

In this section, we will prove that the approximation computed by our numerical method converges to a weak solution of (1.1) as discretization parameters go to zero. But, before we embark on this task, let us first recall the notion of weak solutions associated with (1.1).

3.1. Weak formulation. Due to the presence of the Lagrange multiplier γ, weak solutions are standardly defined using the angular momentum w. Specifically, since $\gamma = |\nabla d|^2 - |d_t|^2$ the energy only provides an L^1 bound on γ. For this reason, the weak formulation of (1.1) is often posed using (1.3) and the following integration by parts formula:

LEMMA 3.1. *For all sufficiently smooth functions (d, ϕ), there holds*

$$\int_\Omega (d \times \Delta d)\phi \, dx = \int_\Omega (\nabla d \times d) : \nabla \phi \, dx.$$

PROOF. To derive the identity, we calculate

$$\begin{aligned} d \times \Delta d &= \begin{pmatrix} d_2 \Delta d_3 - d_3 \Delta d_2 \\ d_3 \Delta d_1 - d_1 \Delta d_3 \\ d_1 \Delta d_2 - d_2 \Delta d_1 \end{pmatrix} = \operatorname{div} \begin{pmatrix} \nabla d_3 \cdot d_2 - \nabla d_2 \cdot d_3 \\ \nabla d_1 \cdot d_3 - \nabla d_3 \cdot d_1 \\ \nabla d_2 \cdot d_1 - \nabla d_1 \cdot d_2 \end{pmatrix} \\ &= -\operatorname{div} (\nabla d \times d) . \end{aligned} \tag{3.1}$$

Multiplying with ϕ and integrating over the domain concludes the proof. □

REMARK 3.2. A discrete version of (3.1) can be readily deduced for the operators Δ_h, ∇_h, and div_h. Indeed, the proof only relies on the identity $\operatorname{div}(\nabla a \; b) = b\Delta a + \nabla a \cdot \nabla b$, which is satisfied by the numerical operators.

The weak formulation of (1.1) is given by the following definition. We refer to [**111**] for more on this formulation and corresponding existence theory.

DEFINITION 3.3. Given initial data d^0, d_t^0, with finite energy

$$E(d^0) := \frac{1}{2} \left(\|d_t^0\|^2_{L^2(\Omega)} + \|\nabla d^0\|^2_{L^2(\Omega)} \right) \leq C,$$

and $|d^0| = 1$ a.e., we call d a weak solution of (1.1) provided:

(1) the energy satisfies

$$\sup_t E(d) \leq E(d^0).$$

(2) the following weak formulation holds

$$\int_0^\infty \int_\Omega -(d_t \times d)\phi_t + (\nabla d \times d) : \nabla \phi \; dxdt = \int_\Omega (d_t^0 \times d^0)\phi(0,\cdot) \; dx, \tag{3.2}$$

for all $\phi \in C_c^\infty([0,\infty) \times \Omega)$.

(3) the initial condition is satisfied, i.e., as $t \to 0$,

$$d \rightharpoonup d^0 \text{ in } W^{1,2}(\Omega), \qquad d_t \rightharpoonup d_t^0 \text{ in } L^2(\Omega)$$

3.2. Main convergence result. Our main result in this section is the following convergence result:

THEOREM 3.4. *Let $\{(d_h, w_h)\}_{h>0}$ be a sequence of numerical approximations obtained using Definition* 2.1 *and* (2.3)*, where $\Delta t = \kappa h$ for some constant $\kappa > 0$. Then, as $h \to 0$, $d_h \to d$ a.e. and in $L^p((0,\infty) \times \Omega)$ for any $p < \infty$, $w_h \overset{\star}{\rightharpoonup} w$ in $L^\infty(0,T; L^2(\Omega))$, where*

$$\begin{aligned} |d| &= 1, \ \text{a.e. in } (0,\infty) \times \Omega, \\ w &= d_t \times d, \ \text{a.e. in } [0,\infty) \times \Omega, \end{aligned}$$

Furthermore, d is a weak solution of the wave map equation (1.1) *in the sense of Definition* 3.3.

To prove this theorem, our starting point is Lemma 2.2 yielding the h-independent bounds:

$$\begin{aligned} D_t d_h &\in_b L^\infty(0,\infty; L^2(\Omega)), \\ \nabla_h d_h &\in_b L^\infty(0,\infty; L^2(\Omega)), \\ w_h &\in_b L^\infty(0,\infty; L^2(\Omega)), \end{aligned}$$

where the $\in_b$ means that the inclusion is independent of h. From these bounds, we can assert the existence of functions d and w, and a subsequence h_j, such that

$$\begin{aligned} w_h &\overset{\star}{\rightharpoonup} w \text{ in } L^\infty(0,\infty; L^2(\Omega)), \\ D_t d_h &\overset{\star}{\rightharpoonup} d_t \text{ in } L^\infty(0,\infty; L^2(\Omega)), \\ \nabla_h d_h &\overset{\star}{\rightharpoonup} \nabla d \text{ in } L^\infty(0,\infty; L^2(\Omega)), \\ d_h &\to d \text{ a.e. and in } L^p((0,\infty) \times \Omega) \text{ for } p < \infty, \end{aligned} \tag{3.3}$$

where the limit d also satisfies the constraint

$$|d(t,x)| = 1 \text{ a.e. in } [0,\infty) \times \Omega.$$

3.3. Proof of Theorem 3.4: For test functions $\varphi, \psi \in C_0^1([0,\infty)\times\Omega;\mathbb{R}^n)$, we denote $\varphi^m(x) := \varphi(t^m, x)$, $\psi^m(x) := \psi(t^m, x)$. Then we dot (2.4) and (2.5) with φ^m, ψ^m, integrate over Ω, and sum over m, to discover

$$\Delta t \sum_{m=0}^{\infty} \int_\Omega \left(D_t d_h^m - d_h^{m+1/2} \times w_h^{m+1/2} \right) \cdot \varphi^m \, dx = 0,$$
$$\Delta t \sum_{m=0}^{\infty} \int_\Omega \left(D_t w_h^m + d_h^{m+1/2} \times \Delta_h d_h^{m+1/2} \right) \cdot \psi^m \, dx = 0.$$

Using Lemma 3.1 (see Remark 3.2) and summation by parts, we deduce that

$$\begin{aligned} \Delta t \sum_{m=0}^{\infty} \int_\Omega \left(-d_h^{m+1} \cdot D_t \varphi^m - \left(d_h^{m+1/2} \times w_h^{m+1/2} \right) \cdot \varphi^m \right) dx - \int_\Omega d_h^0 \cdot \varphi^0 \, dx = 0, \\ \Delta t \sum_{m=0}^{\infty} \int_\Omega \left(-w_h^{m+1} \cdot D_t \psi^m - \left(\nabla_h d_h^{m+1/2} \times d_h^{m+1/2} \right) : \nabla_h \psi^m \right) dx - \int_\Omega w_h^0 \cdot \psi^0 \, dx = 0. \end{aligned} \tag{3.4}$$

We denote

$$\begin{aligned} d_h(t,x) &= d_h^m(x), \quad x \in \Omega, \quad t \in (t^{m-1}, t^m], \\ w_h(t,x) &= w_h^m(x), \quad x \in \Omega, \quad t \in (t^{m-1}, t^m], \\ \overline{d}_h(t,x) &= d_h^{m-1/2}(x), \quad x \in \Omega, \quad t \in (t^{m-1}, t^m], \\ \overline{w}_h(t,x) &= w_h^{m-1/2}(x), \quad x \in \Omega, \quad t \in (t^{m-1}, t^m], \end{aligned}$$

such that (3.4) becomes

$$\begin{aligned} -\int_0^\infty \int_\Omega \left(d_h \cdot D_t \varphi + \left(\overline{d}_h \times \overline{w}_h \right) \cdot \varphi \right) dx \, dt - \int_\Omega d_h^0 \cdot \varphi(0,\cdot) \, dx = 0, \\ -\int_0^\infty \int_\Omega \left(w_h \cdot D_t \psi + \left(\nabla_h \overline{d}_h \times \overline{d}_h \right) : \nabla_h \psi \right) dx - \int_\Omega w_h^0 \cdot \psi(0,\cdot) \, dx = 0. \end{aligned}$$

Now, since $\nabla_h \psi \to \nabla \psi$ a.e. and $(D_t\varphi, D_t\psi) \to (\varphi_t, \psi_t)$ a.e., one may apply the convergence statements (3.3) to discover that the limit (d, w) satisfies

$$\begin{aligned} -\int_0^\infty \int_\Omega \left(d \cdot \varphi_t + (d \times w) \cdot \varphi \right) dx \, dt - \int_\Omega d^0 \cdot \varphi(0,\cdot) \, dx = 0, \\ -\int_0^\infty \int_\Omega \left(w \cdot \psi_t + (\nabla d \times d) : \nabla \psi \right) dx - \int_\Omega (d_t^0 \times d^0) \cdot \psi(0,\cdot) \, dx = 0. \end{aligned} \tag{3.5}$$

It only remains to prove that this formulation is equivalent to (3.2) in Definition 3.3. In practice, this means proving that $w = d_t \times d$ since then the second equation in (3.5) becomes (3.2).

By definition of weak derivatives, the first equation in (3.5) tells us that

$$d_t = d \times w \text{ a.e in } (0,T) \times \Omega.$$

Since $|d| = 1$, this means that

$$w = d_t \times d + (d \cdot w)d. \tag{3.6}$$

However, from the numerical method (2.1)-(2.2), we have that

$$D_t(d_h^m w_h^m) = d_h^{m+1/2} D_t w_h^m + w_h^{m+1/2} D_t d_h^m = 0,$$

and since $d_h^0 \cdot w_h^0 = 0$, this means that

$$d_h \cdot w_h = 0, \text{ in } \Omega \text{ for all } t.$$

Since d_h converges strongly and w_h weakly, we conclude that

$$d \cdot w = 0 \text{ a.e in } [0, \infty) \times \Omega.$$

Then, (3.6) becomes

$$w = d_t \times d,$$

and the proof is complete. □

4. A solution may be obtained fast

The new *angular momentum* method (Definition 2.1) is both nonlinear and implicit. Hence, in practice, finding a solution requires solving a fixed point problem at each time step. In this section, we will construct a fixed point iteration scheme and prove that this scheme provides the desired solution using only $N \log N$ operations.

To find a solution of (2.1)–(2.2), we propose the following iterative scheme:

DEFINITION 4.1. Given $h > 0$, $\Delta t = \kappa h$, and functions (d_h^m, w_h^m) satisfying (2.1)–(2.2), we approximate the next time step (d_h^{m+1}, w_h^{m+1}) to a given tolerance $\epsilon > 0$ by the following procedure: Set

$$(d_h^{m,0}, w_h^{m,0}) = (d_h^m, w_h^m),$$

and iteratively solve $(d_h^{m,s+1}, w_h^{m,s+1})$ satisfying

$$\begin{aligned} \frac{d_h^{m,s+1} - d_h^m}{\Delta t} &= \frac{1}{2}\left(d_h^m + d_h^{m,s+1}\right) \times \frac{1}{2}\left(w_h^m + w_h^{m,s}\right), \\ \frac{w_h^{m,s+1} - w_h^m}{\Delta t} &= \frac{1}{2}\left(\Delta_h d_h^m + \Delta_h d_h^{m,s+1}\right) \times \frac{1}{2}\left(d_h^m + d_h^{m,s+1}\right), \end{aligned} \tag{4.1}$$

until the following stopping criteria is met:

$$\left\|w_h^{m,s+1} - w_h^{m,s}\right\|_{L^2(\Omega)} + \left\|\nabla d_h^{m,s+1} - \nabla d_h^{m,s}\right\|_{L^2(\Omega)} < \epsilon. \tag{4.2}$$

Clearly, if the iteration (4.1) yields a fixed point $w_h^{m,s+1} = w_h^{m,s}$, then $(d_h^{m+1}, w_h^{m+1}) = (d_h^{m,s+1}, w_h^{m,s+1})$ is a solution to the nonlinear scheme (2.1)–(2.2). Moreover, the iteration in (4.1) is put up precisely such that the length is preserved at each iteration:

$$|d_h^{m,s}| = |d_h^m| = 1 \quad \text{in } \Omega.$$

Seen from the practical point of view, the remaining questions are whether the iteration converges or not and, if so, how many iterations that are needed to reach the given tolerance

ϵ. The following theorem provides an answer to these questions and is our main result in this section.

THEOREM 4.2. *Given $h > 0$, $\Delta t = \kappa h$ for a sufficiently small $\kappa > 0$, and a small tolerance $\epsilon > 0$, there is a number of iterations $\bar{s} \in \mathbb{N}_+$, $\bar{s} \leq C|\log \epsilon|$, such that* (4.2) *holds and the error*

$$\left\|w_h^{m+1} - w_h^{m,s}\right\|_{L^2(\Omega)} + \left\|\nabla d_h^{m+1} - \nabla d_h^{m,s}\right\|_{L^2(\Omega)} < \epsilon, \quad \forall s \geq \bar{s}. \tag{4.3}$$

The proof of this theorem will follow as a consequence of the results stated and proved in the remaining parts of this section.

REMARK 4.3. In Theorem 4.2, we need that κ is sufficiently small. Upon inspecting the upcoming proof, one can derive that $\kappa \leq \frac{1}{50}$. However, in practice, it is sufficient to have $\kappa \leq \frac{1}{2}$. In all the examples we tested, the fixed point iteration converged as long as $\kappa \lessapprox 0.7$. For higher CFL-numbers the fixed point iteration did not converge anymore in some time steps. This is the only instance at which we need to require anything on κ.

As an immediate corollary of Theorem 4.2, we have that a desired solution may be computed in $N \log N$ operations:

COROLLARY 4.4. *For a given tolerance $\epsilon = N^{-\alpha}$, the functions $(d_h^{m,\bar{s}}, w_h^{m,\bar{s}})$ in Theorem* 4.2 *may be computed using only $\mathcal{O}(N \log N)$ operations.*

PROOF. Since each iteration requires N operations and we need $\mathcal{O}(|\log \epsilon|)$ iterations, we get a total of $\mathcal{O}(N|\log \epsilon|)$ iterations and the proof follows by inserting $\epsilon = N^{-\alpha}$. □

Another consequence of Theorem 4.2 is that the energy at the stopping time $\bar{s}$ is almost conserved:

COROLLARY 4.5. *Under the conditions of Theorem* 4.2*,*

$$E_m^{\bar{s}} := \frac{1}{2}\left(\left\|w_h^{m,\bar{s}}\right\|_{L^2(\Omega)}^2 + \left\|\nabla_h d_h^{m,\bar{s}}\right\|_{L^2(\Omega)}^2\right) = E_0 + \mathcal{O}(\epsilon).$$

PROOF. By multiplying the first equation in (4.1) with $-\frac{1}{2}\Delta_h(d_h^{m,s+1} + d_h^m)$, the second equation with $\frac{1}{2}(w_h^{m,s} + w_h^m)$, and integrating by parts, we obtain that

$$\begin{aligned}
&\frac{1}{2}\left(\left\|w_h^{m,s+1}\right\|_{L^2(\Omega)}^2 + \left\|\nabla_h d_h^{m,s+1}\right\|_{L^2(\Omega)}^2\right) \\
&\quad = E_m + \frac{1}{2}\int_\Omega (w_h^{m,s+1} - w_h^m)(w_h^{m,s+1} - w_h^{m,s})\, dx \\
&\quad = E_0 + \frac{1}{2}\int_\Omega (w_h^{m,s+1} - w_h^m)(w_h^{m,s+1} - w_h^{m,s})\, dx,
\end{aligned}$$

where the last inequality follows from Lemma 2.2.

Finally, we assume that $s > \bar{s}$ such that both (4.2) and (4.3) holds. The Cauchy-Schwarz inequality then provides the estimate

$$\begin{aligned}&\frac{1}{2}\int_\Omega (w_h^{m,s+1} - w_h^m)(w_h^{m,s+1} - w_h^{m,s})\, dx \\ &\leq \frac{1}{2}\|w_h^{m,s+1} - w_h^m\|_{L^2(\Omega)}\|w_h^{m,s+1} - w_h^{m,s}\|_{L^2(\Omega)} \leq C\frac{\epsilon}{2},\end{aligned}$$

which brings the proof to an end. □

REMARK 4.6. In practice, the (d_h^m, w_h^m) appearing in the fixed point scheme (4.1) would be the approximation coming from the previous time step. In this case Corollary 4.5 tells us that the "error" will be summed and thus

$$E_m^{\bar{s}} = E_0 + \mathcal{O}(m\epsilon).$$

4.1. The fixed point map F_m. To prove Theorem 4.2, it will be convenient to write the fixed point iteration in terms of a map. To define this map, we first notice that (2.1), can be rewritten as

$$d_{\underline{i}}^{m+1} = V(w_{\underline{i}}^{m+1/2})\, d_{\underline{i}}^m \tag{4.4}$$

where $V = V(w)$ is the following matrix

$$V(w) = \frac{1}{1 + \frac{\Delta t^2}{4}|w|^2}\left(\left(1 - \frac{\Delta t^2}{4}|w|^2\right)\mathbb{1} + \frac{\Delta t^2}{2}(w \otimes w) + \Delta t Q(w)\right), \tag{4.5}$$

and $Q(w)$ is defined as

$$Q(w) = \begin{pmatrix} 0 & w^{(3)} & -w^{(2)} \\ -w^{(3)} & 0 & w^{(1)} \\ w^{(2)} & -w^{(1)} & 0 \end{pmatrix}.$$

In particular, $Q(\cdot)$ is such that

$$Q(w)v = v \times w$$

for any vector $v \in \mathbb{R}^3$. Note that V is an orthogonal matrix, and therefore, independently of w,

$$|V(w)v|^2 = |v|^2$$

for any $v \in \mathbb{R}^3$.

To prove the theorem, we will demonstrate that w_h^{m+1} is the fixed point of a contractive mapping F_m which is defined as follows:

DEFINITION 4.7 (The mapping F_m). For a piecewise constant function u_h on Ω,

$$u_h(x) = u_{\underline{i}}, \quad x \in E_{\underline{i}},\ \underline{i} \in \mathcal{I}_N, \tag{4.6}$$

for some $\{u_{\underline{i}}\}_{\underline{i} \in \mathcal{I}_N}$, we define the piecewise constant function $v_h := F_m(u_h)$ by

$$v_h(x) = v_{\underline{i}}, \quad x \in E_{\underline{i}},\ \underline{i} \in \mathcal{I}_N,$$

where $v_{\underline{i}}$, $\underline{i} \in \mathcal{I}_N$ is given by

$$\begin{aligned} v_{\underline{i}} &= w_{\underline{i}}^m + \Delta t \big[\Delta_h \left(\overline{V}(\overline{u}_{\underline{i}}) d_{\underline{i}}^m\right)\big] \times \left(\overline{V}(\overline{u}_{\underline{i}}) d_{\underline{i}}^m\right) \\ \overline{u}_{\underline{i}} &= \frac{w_{\underline{i}}^m + u_{\underline{i}}}{2}, \quad \underline{i} \in \mathcal{I}_N, \end{aligned} \tag{4.7}$$

and $\overline{V} := (\mathbb{1} + V)/2$.

A fixed point $v_h = v_h^*$ of F_m, will be a solution to (2.2) and d_h^* defined as a piecewise constant interpolation of $d_{\underline{i}}^* = V(\overline{v}_{\underline{i}}^*)\, d_{\underline{i}}^m$, $\underline{i} \in \mathcal{I}_N$, will be a solution to (2.1).

4.2. The map F_m is a contraction. We now proceed to proving that the mapping F_m is a contraction.

LEMMA 4.8. *The mapping F_m defined by (4.7) is a contraction in the $L^2(\Omega)$-norm if $\Delta t \le \kappa h$ for a constant κ sufficiently small, that is,*

$$\|F_m(u_{1,h}) - F_m(u_{2,h})\|_{L^2(\Omega)} \le q \|u_{1,h} - u_{2,h}\|_{L^2(\Omega)}$$

for some $q < 1$ for any two piecewise constant functions $u_{1,h}$, $u_{2,h}$ on Ω defined as in (4.6). In particular, by Banach's fixed point theorem, this implies that the mapping F_m has a unique fixed point.

PROOF. For the ease of notation, we will omit the indices $\underline{i}$, m and h and write w, d, F, u_1, u_2 for w_h^m, d_h^m, F_m, $u_{1,h}$, $u_{2,h}$, respectively. Moreover, we denote $y_1 := F(u_1)$ and $y_2 := F(u_2)$ and $\overline{u}_j := (w + u_j)/2$, $j = 1, 2$, such that

$$y_j = w + \Delta t \operatorname{div}_h \big[\nabla_h \left(\overline{V}\left(\overline{u}_j\right) d\right) \times \overline{V}\left(\overline{u}_j\right) d\big], \quad j = 1, 2.$$

Then, using the inverse inequality,

$$\begin{aligned} \|y_1 - y_2\| &= \Delta t \big\| \operatorname{div}_h \big[\nabla_h \left(\overline{V}(\overline{u}_1) d\right) \times \overline{V}(\overline{u}_1) - \nabla_h \left(\overline{V}(\overline{u}_2) d\right) \times \overline{V}(\overline{u}_2) d\big] \big\|_{L^2(\Omega)} \\ &\le C \frac{\Delta t}{h} \big\| \nabla_h \left(\overline{V}(\overline{u}_1) d\right) \times \overline{V}(\overline{u}_1) - \nabla_h \left(\overline{V}(\overline{u}_2) d\right) \times \overline{V}(\overline{u}_2) d \big\|_{L^2(\Omega)} \\ &\le C \frac{\Delta t}{h} \Big(\big\| \nabla_h \left([V(\overline{u}_1) - V(\overline{u}_2)] d\right) \times \overline{V}(\overline{u}_1) d \big\|_{L^2(\Omega)} \\ &\qquad + \big\| \nabla_h \left(\overline{V}(\overline{u}_2) d\right) \times [V(\overline{u}_1) - V(\overline{u}_2)] d \big\|_{L^2(\Omega)} \Big) \\ &\le C \frac{\Delta t}{h^2} \big\| [V(\overline{u}_1) - V(\overline{u}_2)] d \big\|_{L^2(\Omega)}, \end{aligned} \tag{4.8}$$

using that $|\overline{V}(\overline{u}_j) d| \le 1$ for the last inequality. We split $\|[V(\overline{u}_1) - V(\overline{u}_2)] d\|_{L^2(\Omega)}$ using (4.5),

$$\begin{aligned} &\big\| [V(\overline{u}_1) - V(\overline{u}_2)] d \big\|_{L^2(\Omega)} \\ &\quad \le \left\| \left[\frac{1 - \frac{\Delta t^2}{4} |\overline{u}_1|^2}{1 + \frac{\Delta t^2}{4} |\overline{u}_1|^2} - \frac{1 - \frac{\Delta t^2}{4} |\overline{u}_2|^2}{1 + \frac{\Delta t^2}{4} |\overline{u}_2|^2} \right] d \right\|_{L^2(\Omega)} \end{aligned}$$

$$
\begin{aligned}
&+ \left\| \left[\frac{\Delta t^2}{2+\frac{\Delta t^2}{2}|\overline{u}_1|^2}(\overline{u}_1 \otimes \overline{u}_1) - \frac{\Delta t^2}{2+\frac{\Delta t^2}{2}|\overline{u}_2|^2}(\overline{u}_2 \otimes \overline{u}_2) \right] d \right\|_{L^2(\Omega)} \\
&+ \left\| \left[\frac{\Delta t}{1+\frac{\Delta t^2}{4}|\overline{u}_1|^2} Q(\overline{u}_1) - \frac{\Delta t}{1+\frac{\Delta t^2}{4}|\overline{u}_2|^2} Q(\overline{u}_2) \right] d \right\|_{L^2(\Omega)} \\
&=: \mathrm{I} + \mathrm{II} + \mathrm{III}.
\end{aligned}
$$

For the I term, we apply the Cauchy-Schwarz inequality to discover that

$$
\begin{aligned}
\mathrm{I} &= \left\| \frac{\frac{\Delta t^2}{2}\left(|\overline{u}_1|^2 - |\overline{u}_2|^2\right)}{\left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right)\left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right)} d \right\|_{L^2(\Omega)} \\
&\leq \Delta t \left\| \frac{1+\frac{\Delta t^2}{4}|\overline{u}_1|^2 + \frac{\Delta t^2}{4}|\overline{u}_2|^2}{\left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right)\left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right)} |\overline{u}_1 - \overline{u}_2| d \right\|_{L^2(\Omega)} \\
&\leq \Delta t \|u_1 - u_2\|_{L^2(\Omega)},
\end{aligned} \tag{4.9}
$$

where we have used $|d| = 1$ to conclude the last inequality.

To bound the II term, we first note that

$$
\begin{aligned}
\mathrm{II} &= \frac{\Delta t^2}{2} \left\| \frac{\left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right)(\overline{u}_1 \otimes \overline{u}_1) - \left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right)(\overline{u}_2 \otimes \overline{u}_2)}{\left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right)\left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right)} d \right\|_{L^2(\Omega)} \\
&= \frac{\Delta t^2}{2} \Bigg(\int \alpha \sum_{i=1}^{3} \Bigg(\sum_{j=1}^{3} \Bigg[\left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right) \overline{u}_1^{(i)} \overline{u}_1^{(j)} \\
&\qquad - \left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right) \overline{u}_2^{(i)} \overline{u}_2^{(j)} \Bigg] d^{(j)} \Bigg)^2 dx \Bigg)^{\frac{1}{2}}
\end{aligned}
$$

where

$$
\alpha = \frac{1}{\left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right)\left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right)}.
$$

Since $|d^{(j)}| \leq 1$, $j = 1, 2, 3$,

$$
\begin{aligned}
\mathrm{II} &\leq \frac{\Delta t^2}{2} \left(\int \alpha \sum_{i=1}^{3} \left(\sum_{j=1}^{3} \left| \left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right) \overline{u}_1^{(i)} \overline{u}_1^{(j)} - \left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right) \overline{u}_2^{(i)} \overline{u}_2^{(j)} \right| \right)^2 dx \right)^{\frac{1}{2}} \\
&\leq \frac{3\Delta t^2}{2} \sum_{i,j=1}^{3} \left\| \alpha \left(\left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right) \overline{u}_1^{(i)} \overline{u}_1^{(j)} - \left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right) \overline{u}_2^{(i)} \overline{u}_2^{(j)} \right) \right\|_{L^2(\Omega)}.
\end{aligned}
$$

We consider one of the summands:

$$
\begin{aligned}
&\left\| \alpha \left(\left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right) \overline{u}_1^{(i)} \overline{u}_1^{(j)} - \left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right) \overline{u}_2^{(i)} \overline{u}_2^{(j)} \right) \right\|_{L^2(\Omega)} \\
&\leq \left\| \alpha (\overline{u}_1^{(i)} \overline{u}_1^{(j)} - \overline{u}_2^{(i)} \overline{u}_2^{(j)}) \right\|_{L^2(\Omega)} + \frac{\Delta t^2}{4} \left\| \alpha \left(|\overline{u}_2|^2 \overline{u}_1^{(i)} \overline{u}_1^{(j)} - |\overline{u}_1|^2 \overline{u}_2^{(i)} \overline{u}_2^{(j)} \right) \right\|_{L^2(\Omega)}
\end{aligned}
$$

$$=: \mathrm{II}_1 + \mathrm{II}_2.$$

By adding and subtracting, and applying the Cauchy-Schwarz inequality, we deduce the following bound for the first term,

$$\begin{aligned}
\mathrm{II}_1 &= \frac{1}{\Delta t}\big\|\alpha(\Delta t\,\overline{u}_1^{(i)}(u_1^{(j)} - u_2^{(j)}) + (u_1^{(i)} - u_2^{(i)})\Delta t\,\overline{u}_2^{(j)})\big\|_{L^2(\Omega)} \\
&\leq \frac{1}{\Delta t}\Big\{\big\|\alpha\Delta t\overline{u}_1^{(i)}(u_1^{(j)} - u_2^{(j)})\big\|_{L^2(\Omega)} + \big\|\alpha(u_1^{(i)} - u_2^{(i)})\Delta t\overline{u}_2^{(j)}\big\|_{L^2(\Omega)}\Big\} \\
&\leq \frac{1}{2\Delta t}\Big\{\big\|\alpha(1+\Delta t^2(\overline{u}_1^{(i)})^2)(u_1^{(j)} - u_2^{(j)})\big\|_{L^2(\Omega)} \\
&\qquad + \big\|\alpha(u_1^{(i)} - u_2^{(i)})(1+\Delta t^2(\overline{u}_2^{(j)})^2)\big\|_{L^2(\Omega)}\Big\} \\
&\leq \frac{4}{\Delta t}\|u_1 - u_2\|_{L^2(\Omega)}, \qquad (4.10)
\end{aligned}$$

where the last inequality follows by inserting the definition of α.

Term II_2 may be written as

$$\mathrm{II}_2 = \frac{\Delta t^2}{4}\left\|\alpha\sum_{k=1}^{3}\left((\overline{u}_2^{(k)})^2\overline{u}_1^{(i)}\overline{u}_1^{(j)} - (\overline{u}_1^{(k)})^2\overline{u}_2^{(i)}\overline{u}_2^{(j)}\right)\right\|_{L^2(\Omega)}.$$

We consider one of the terms in the sum. Note that if $i = j$, the term where $i = j = k$ cancels, hence we can assume without loss of generality that $i \neq k$. By adding and subtracting, we rewrite one of the terms in II_2 as follows

$$\begin{aligned}
&(\overline{u}_2^{(k)})^2\overline{u}_1^{(i)}\overline{u}_1^{(j)} - (\overline{u}_1^{(k)})^2\overline{u}_2^{(i)}\overline{u}_2^{(j)} \\
&\quad = \overline{u}_1^{(i)}\overline{u}_1^{(j)}\overline{u}_2^{(k)}(u_2^{(k)} - u_1^{(k)}) + \overline{u}_1^{(k)}\overline{u}_2^{(k)}\overline{u}_1^{(j)}(u_1^{(i)} - u_2^{(i)}) \\
&\qquad + \overline{u}_1^{(k)}\overline{u}_2^{(k)}\overline{u}_2^{(i)}(u_1^{(j)} - u_2^{(j)}) + \overline{u}_2^{(j)}\overline{u}_2^{(i)}\overline{u}_1^{(k)}(u_2^{(k)} - u_1^{(k)}).
\end{aligned}$$

Next, we apply Young's inequality to the previous identity giving

$$\begin{aligned}
&\big|(\overline{u}_2^{(k)})^2\overline{u}_1^{(i)}\overline{u}_1^{(j)} - (\overline{u}_1^{(k)})^2\overline{u}_2^{(i)}\overline{u}_2^{(j)}\big| \\
&\quad \leq \frac{1}{\Delta t}\left(|\overline{u}_1|^2 + |\overline{u}_2|^2 + \frac{\Delta t^2}{4}|\overline{u}_1|^2|\overline{u}_2|^2\right) \times \left(|u_1^{(i)} - u_2^{(i)}| + |u_1^{(j)} - u_2^{(j)}| + |u_1^{(k)} - u_2^{(k)}|\right) \\
&\quad = \frac{4}{(\Delta t)^3}\left(\frac{1}{\alpha} - 1\right) \times \left(|u_1^{(i)} - u_2^{(i)}| + |u_1^{(j)} - u_2^{(j)}| + |u_1^{(k)} - u_2^{(k)}|\right).
\end{aligned}$$

As a consequence, we conclude that

$$\begin{aligned}
\mathrm{II}_2 &= \frac{\Delta t^2}{4}\left\|\alpha\sum_{k=1}^{3}\left((\overline{u}_2^{(k)})^2\overline{u}_1^{(i)}\overline{u}_1^{(j)} - (\overline{u}_1^{(k)})^2\overline{u}_2^{(i)}\overline{u}_2^{(j)}\right)\right\|_{L^2(\Omega)} \\
&\leq \frac{1}{\Delta t}\|\overline{u}_1 - \overline{u}_2\|_{L^2(\Omega)}. \qquad (4.11)
\end{aligned}$$

From (4.10) and (4.11), we have that

$$\mathrm{II} \leq 25\Delta t\|\overline{u}_1 - \overline{u}_2\|_{L^2(\Omega)}. \qquad (4.12)$$

The final term III can be bounded as follows

$$
\begin{aligned}
\mathrm{III} &= \Delta t \left\| \frac{\left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right) Q(\overline{u}_1) - \left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right) Q(\overline{u}_2)}{\left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right)\left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right)} d \right\|_{L^2(\Omega)} \\
&= \Delta t \left\| \frac{d \times \left[\left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right)\overline{u}_1 - \left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right)\overline{u}_2\right]}{\left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right)\left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right)} \right\|_{L^2(\Omega)} \\
&\leq \Delta t \left\| \frac{u_1 - u_2 + \frac{\Delta t^2}{4}\left(|\overline{u}_2|^2\overline{u}_1 - |\overline{u}_1|^2\overline{u}_2\right)}{\left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right)\left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right)} \right\|_{L^2(\Omega)} \\
&\leq \Delta t \left\| \frac{\left[1+\frac{\Delta t^2}{8}\left(|\overline{u}_1|^2+|\overline{u}_2|^2\right)\right](u_1-u_2)}{\left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right)\left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right)} + \frac{\frac{\Delta t^2}{8}\left(|\overline{u}_1+\overline{u}_2|^2\,|\overline{u}_1-\overline{u}_2|\right)}{\left(1+\frac{\Delta t^2}{4}|\overline{u}_1|^2\right)\left(1+\frac{\Delta t^2}{4}|\overline{u}_2|^2\right)} \right\|_{L^2(\Omega)} \\
&\leq 2\Delta t \, \|u_1 - u_2\|_{L^2(\Omega)}.
\end{aligned}
\tag{4.13}
$$

Summing up (4.8), (4.9), (4.12) and (4.13), we find

$$\|y_1 - y_2\|_{L^2(\Omega)} \leq C\frac{\Delta t^2}{h^2}\|\overline{u}_1 - \overline{u}_2\|_{L^2(\Omega)},$$

and hence the map F_m is a contraction as long as $\Delta t \leq \kappa h$ for a constant $\kappa > 0$ small enough. This concludes the proof. □

The previous lemma immediately provides the existence of a unique solution to (2.1)–(2.2).

COROLLARY 4.9. *Given a previous time step (d_h^m, w_h^m), there exists a unique numerical solution (d_h^{m+1}, w_h^{m+1}) to the numerical method given in Definition* 2.1.

4.3. Proof of Theorem 4.2. Using the previous lemma, we can now prove that the fixed point iteration in Definition 4.1 will converge to the correct solution.

Theorem 4.2 is an immediate consequence of the following lemma.

LEMMA 4.10. *Given any $\epsilon_0 > 0$, there is a number of iterations $\bar{s} \in \mathbb{N}_+$ in Definition* 4.1 *with $\bar{s} \leq C|\log \epsilon_0|$ such that* (4.2) *holds with $\epsilon = \epsilon_0$ and*

$$\|w_h^{m+1} - w_h^{m,\bar{s}}\|_{L^2(\Omega)} + \|\nabla d_h^{m+1} - \nabla d_h^{m,\bar{s}}\|_{L^2(\Omega)} \leq \epsilon_0.$$

PROOF. We again omit writing the indices h and $\underline{i}$ and denote $w^{m,0} := w^m$, $w^{m,s} := F_m(w^{m,s-1})$ for $s \geq 1$. Now, since F_m is a contraction with 'Lipschitz' constant $q < 1$ and $F_m(w^{m+1}) = w^{m+1}$,

$$
\begin{aligned}
\|w^{m+1} - w^{m,s}\|_{L^2(\Omega)} &= \|F(w^{m+1}) - F(w^{m,s-1})\|_{L^2(\Omega)} \\
&\leq q\|w^{m+1} - w^{m,s-1}\|_{L^2(\Omega)} \\
&\leq q^s\|w^{m+1} - w^m\|_{L^2(\Omega)}.
\end{aligned}
$$

Thus, it follows from the energy estimate,

$$\|w^{m+1} - w^{m,s}\|_{L^2} \leq 2q^s E_0. \tag{4.14}$$

Moreover, we note that it follows by the inverse inequality, (4.4), (4.5) and (4.9), (4.12) and (4.13),

$$\begin{aligned}\|\nabla_h d_h^{m,s} - \nabla_h d_h^{m+1}\|_{L^2(\Omega)} &= \big\|\nabla_h\big[V((w_h^m + w_h^{m,s-1})/2) - V(w_h^{m+1/2})\big]d_h^m\big\|_{L^2(\Omega)} \\ &\leq \frac{C}{h}\Big\|\Big(V((w_h^m + w_h^{m,s-1})/2) - V(w_h^{m+1/2})\Big)\, d_h^m\Big\|_{L^2(\Omega)} \\ &\leq \frac{C\Delta t}{h}\|w_h^{m,s-1} - w_h^{m+1}\|_{L^2(\Omega)} \leq 2\,C q^{s-1} E_0, \end{aligned} \tag{4.15}$$

where the last inequality follows from the CFL-condition and (4.14). Hence, by using the triangle inequality,

$$\|w^{m,s+1} - w^{m,s}\|_{L^2(\Omega)} \leq \|w^{m,s+1} - w^{m+1}\|_{L^2(\Omega)} + \|w^{m+1} - w^{m,s}\|_{L^2(\Omega)} \leq 4\, q^s E_0,$$

and for this reason also

$$\|\nabla_h d^{m,s+1} - \nabla_h d^{m,s}\|_{L^2(\Omega)} \leq 4\,C\, q^{s-1} E_0,$$

which implies that the fixed point iteration converges. That is, the stopping criteria (4.2) is met once s is high enough to satisfy

$$4(Cq^{-1}+1)\, q^s E_0 < \epsilon_0 \quad \Rightarrow \quad s > \frac{\log\left(\frac{4(Cq^{-1}+1)E_0}{\epsilon_0}\right)}{\log\left(\frac{1}{q}\right)}.$$

From (4.14) and (4.15), it is clear that this s also satisfies

$$\big\|w^{m+1} - w^{m,s}\big\|_{L^2(\Omega)} + \big\|\nabla_h d_h^{m,s} - \nabla_h d_h^{m+1}\big\|_{L^2(\Omega)} < \epsilon_0.$$

This concludes the proof. □

5. Numerical results

In this final section, we shall report on some numerical experiments with the new angular momentum method. We shall consider two cases. In the first case, we will explore the rate of convergence of the method. In the second case, we will check whether the method predicts blow-up of the gradient for initial data where this is known to be the case.

5.1. Convergence test. It is a non-trivial task to find analytical solutions of the wave map equation (1.1) in $3D$. In $2D$ however, the dynamics of the wave map equation may be totally described by the linear wave equation. Specifically, upon introducing an angle $\vartheta(t,x)$ and writing

$$d(t,x) = \begin{pmatrix}\cos\vartheta \\ \sin\vartheta,\end{pmatrix},$$

one easily derives that ϑ evolves according to the linear wave equation

$$\vartheta_{tt} - \Delta\vartheta = 0. \tag{5.1}$$

Hence, in the $2D$-case, we can compute analytical solutions using d'Alembert's formula. In particular, (5.1) has solutions of the form

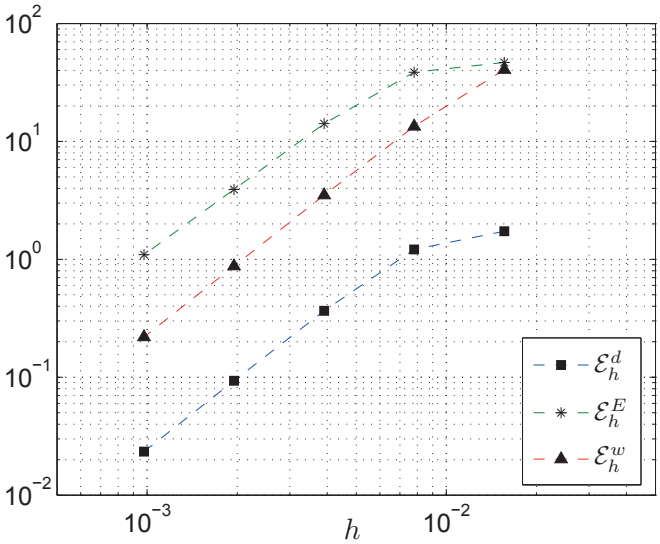

FIGURE 1. The errors $\mathcal{E}_h^d$, $\mathcal{E}_h^w$ and $\mathcal{E}_h^E$ for the approximations to (1.1) for a solution of the form (5.2) at time $T = 20$ for $h = 2^{-j}$, $j = 6, \dots, 10$.

$$
\begin{aligned}
\vartheta(t,x,y) = \sum_{j=-J}^{J} \big\{ & a_j^+ \sin(2\pi j(\sqrt{2}t + (x+y))) + a_j^- \sin(2\pi j(\sqrt{2}t - (x+y))) \\
& + b_j^+ \cos(2\pi j(\sqrt{2}t + (x+y))) + b_j^- \cos(2\pi j(\sqrt{2}t - (x+y))) \\
& + c_j^+ \sin(2\pi j(\sqrt{2}t + (x-y))) + c_j^- \sin(2\pi j(\sqrt{2}t - (x-y))) \\
& + d_j^+ \cos(2\pi j(\sqrt{2}t + (x-y))) + d_j^- \cos(2\pi j(\sqrt{2}t - (x-y))) \big\},
\end{aligned} \tag{5.2}
$$

for $J \geq 0$. In Figure 1, we have plotted the errors between the approximations d_h and d, w_h and w respectively, in the L^2-norm and in the energy norm, that is,

$$
\begin{aligned}
\mathcal{E}_h^d &:= \sup_m \|d(t^m, \cdot) - d_h^m\|_{L^2(\Omega)} \\
\mathcal{E}_h^w &:= \sup_m \|w(t^m, \cdot) - w_h^m\|_{L^2(\Omega)} \\
\mathcal{E}_h^E &:= \sup_m \sqrt{\|\nabla d(t^m, \cdot) - \nabla_h d_h^m\|_{L^2(\Omega)}^2 + \|d_t(t^m, \cdot) - D_t d_h^m\|_{L^2(\Omega)}^2}
\end{aligned}
$$

for $a_1^+ = a_1^- = 1/4$, $a_2^+ = a_2^- = 1/10$, $b_1^+ = -b_1^- = -2$, $b_2^+ = -b_2^- = 1/100$, $T = 20$, $\Delta t = 0.5h$, and $\Omega = \mathbb{T}^2$. Moreover, $h_j = 2^{-j}$, $j = 6, \dots, 10$ for tolerance $\epsilon_0 = h^2$. We observe a rate of convergence of almost 2 for $\mathcal{E}_h^d$ and $\mathcal{E}_h^w$ and about 1.8 for $\mathcal{E}_h^E$ (Table 1).

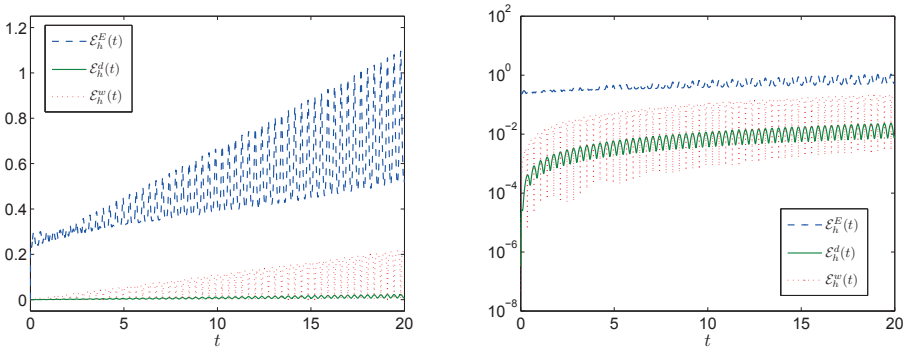

FIGURE 2. The evolution of the errors $\mathcal{E}_h^d(t)$, $\mathcal{E}_h^w(t)$ and $\mathcal{E}_h^E(t)$ versus time for the approximations to (1.1) for a solution of the form (5.2) for $h = 2^{-10}$. Left: Real error, right: Error in log-scale.

Other choices of ϵ_0 such as $h^{3/2}$ or h^3 gave similar results. In Figure 2, the evolution of

h	$\mathcal{E}_h^d$	$\mathcal{E}_h^E$	$\mathcal{E}_h^w$
2^{-6}	1.731	46.78	40.58
2^{-7}	1.213	38.64	13.42
2^{-8}	0.366	14.15	3.499
2^{-9}	0.093	3.915	0.877
2^{-10}	0.023	1.096	0.219
Rate	1.56	1.35	1.88

TABLE 1. Errors for different mesh resolutions for (1.1), (5.2) at time $T = 20$, $\Delta t = 0.5h$, and average rate for grid sizes 2^{-6} to 2^{-10}.

the errors $\mathcal{E}_h^\alpha(t)$, $\alpha \in \{d, w, E\}$, where $\mathcal{E}_h^d(t) := \|d(t,\cdot) - d_h(t,\cdot)\|_{L^2(\Omega)}$, and the other two defined in a similar way, for $h = 2^{-10}$ versus time is shown. It appears that after an initial exponential increase, the error increases linearly with time.

5.2. Initial data developing singularities. In our second experiment, we compare the approximations computed by (2.1)–(2.2) to those obtained with the algorithms from [**9**]. Specifically, we compute approximations for the initial data,

$$d^0(x,y) = \begin{cases} (0,0,-1)^T, & r \geq 1/2, \\ (2xa, 2ya, a^2 - r^2)^T/(a^2+r^2), & r < 1/2, \end{cases}, \quad w^0(x,y) \equiv 0, \tag{5.3}$$

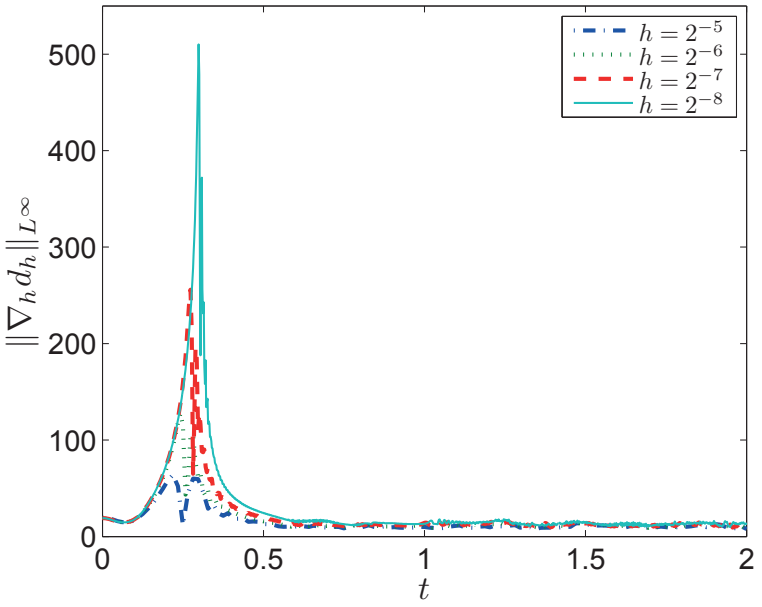

FIGURE 3. The evolution of $\|\nabla_h d_h\|_{L^\infty(\Omega)}$ versus time for initial data (5.3).

on $D = [-0.5, 0.5]^2$, where $r := \sqrt{x^2 + y^2}$ and $a(r) = (1 - 2r)^4$ up to time $T = 2$, with CFL-condition $\Delta t = 0.5h$, $h = 2^{-j}$ for $j = 5, 6, 7, 8$, and tolerance $\epsilon_0 = h^2$. As in [**9**], we observe a blow-up of the gradient ∇d in the L^∞-norm around time $T = 0.3$, cf. Figure 3.

The approximation E_m^s of the discrete energy (2.8) is close to being preserved, as we see in Figure 5, left hand side. In the same figure, on the right hand side, we have plotted the quantity

$$H_m := \frac{1}{2} \int_\Omega |D_t d_h^m|^2 + |\nabla_h d_h^m|^2 \, dx,$$

which is not conserved by our scheme, but upper bounded. Indeed, we calculate

$$\begin{aligned} |D_t d_h^m|^2 &= |d_h^{m+1/2} \times w_h^{m+1/2}|^2 \\ &= |d_h^{m+1/2}|^2 |w_h^{m+1/2}|^2 - (d_h^{m+1/2} \cdot w_h^{m+1/2})^2 \\ &\leq |w_h^{m+1/2}|^2 \\ &\leq \frac{1}{2}(|w_h^m|^2 + |w_h^{m+1}|^2), \end{aligned}$$

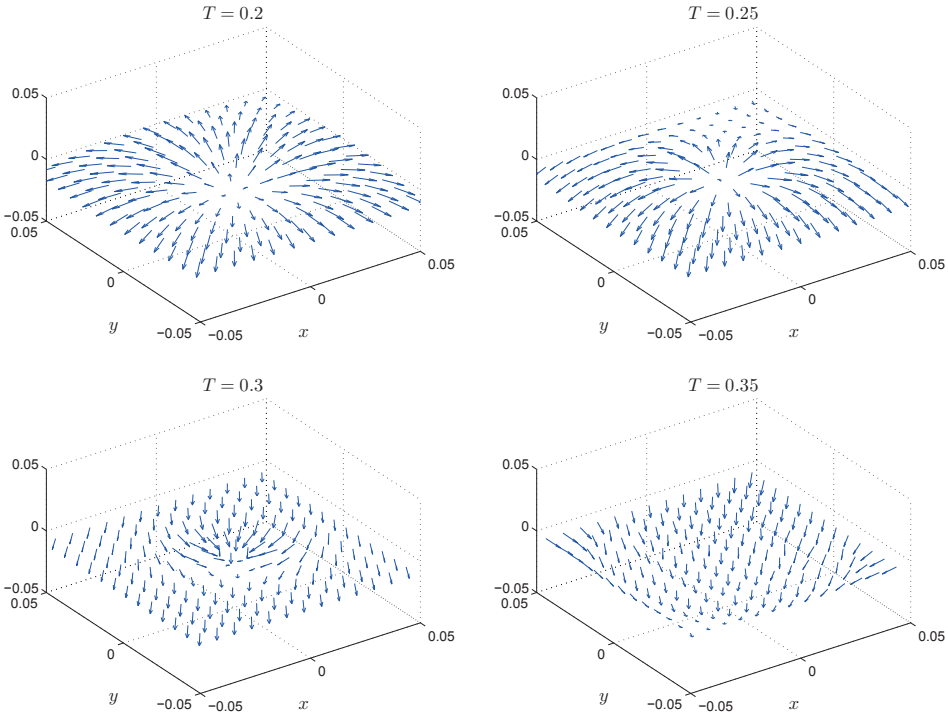

FIGURE 4. The approximation by (2.1)–(2.2) for initial data (5.3) in a neighborhood of the origin before and after blow-up time on a grid with $h = 2^{-7}$.

and hence

$$H_m \leq \frac{1}{2}\int_\Omega \frac{1}{2}(|w_h^m|^2 + |w_h^{m+1}|^2 + |\nabla_h d_h^m|^2\,dx \leq E_m + \frac{1}{2}E_{m+1} = \frac{3}{2}E_0.$$

At the time of singularity formation, there appears to be a rapid transition of energy in ∇d to energy in the angular momentum w, which causes the quantity H_m to overshoot as its definition is based on the approximation at two different time steps, i.e., $D_t d^m$ involves the approximations at time t^m and t^{m+1} whereas ∇d^m is defined at time t^m. We also note that the formation of the singularity causes a loss in H_m, whereas E_m is preserved as predicted by Lemma 2.2. Furthermore, we observe some larger oscillations around the time of blow-up of $\nabla_h d_h$ in the L^∞-norm.

In Figure 4 the approximation of (1.1), (5.3) in a neighborhood of the origin near blow-up time is shown. We observe that the third component of d first switches sign away from the origin and then closer to it, which seems to cause the blow-up in the gradient of d.

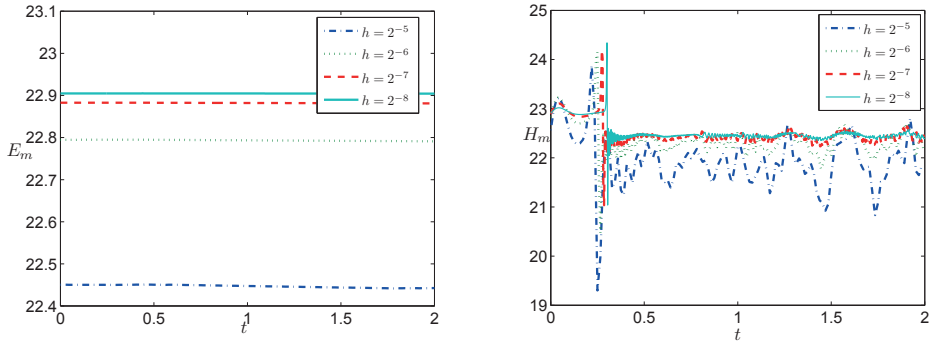

FIGURE 5. The evolution of E_m and H_m versus time for initial data (5.3).

PAPER 2

Analysis and Numerical Approximation of Brinkman Regularization of Two-Phase Flows in Porous Media

Joint work with Giuseppe M. Coclite, Siddhartha Mishra and Nils Henrik Risebro

Abstract. We consider a system of nonlinear partial differential equations that arises in the modeling of two-phase flows in a porous medium. The phase velocities are modeled using a Brinkman regularization of the classical Darcy's law. We propose a notion of weak solution for these equations and prove existence of these solutions. An efficient finite difference scheme is proposed and is shown to converge to the weak solutions of this system. The Darcy limit of the Brinkman regularization is studied numerically using the convergent finite difference scheme in two space dimensions as well as using both analytical and numerical tools in one space dimension. The results suggest that the Brinkman regularization may not approximate the accepted entropy solutions of the Darcy model and raise fundamental questions about the use of Brinkman type models in two-phase flows.

1. The two-phase flow problem

The mathematical description of multi-phase flow in porous media includes a multitude of interesting mathematical models. Interesting both in their own and because they are important for practical simulation of such flows. Perhaps the most prototypical, and also one of the simplest, of such models, describes the flow of two phases, say oil and water in a porous medium. Here the unknowns are the phase saturations s_w and s_o representing the volume fractions of the aqueous and oleic phase respectively. We have the identity:

$$s_w + s_o \equiv 1. \tag{1.1}$$

Hence, we can describe the dynamics in terms of the saturation of either of the two phases. We denote the water saturation by $s_w = s$ in the discussion below. Assuming a constant porosity ($\phi \equiv 1$), mass conservation of the two phases is described by, see [**5**],

$$\partial_t s_r + \mathrm{div}_{\mathbf{x}}(\mathbf{v}_r) = 0, \quad r \in \{w, o\}. \tag{1.2}$$

Here, the phase velocities are denoted by $\mathbf{v}_w$ and $\mathbf{v}_o$ respectively. In view of the identity (1.1), the two phase velocities can be summed up to yield the *incompressibility* condition,

$$\mathrm{div}_{\mathbf{x}}(\mathbf{v}) = 0, \quad \mathbf{v} = \mathbf{v}_w + \mathbf{v}_o. \tag{1.3}$$

The variable $\mathbf{v}$ is called the total velocity.

The phase velocities in a homogeneous isotropic medium are commonly described by Darcy's law [**5, 100**]:

$$\mathbf{v}_r = -\lambda_r \nabla_{\mathbf{x}} p_r, \quad r \in \{w, o\}. \tag{1.4}$$

The quantity $\lambda_r = \lambda_r(s_r)$ is the phase mobility and p_r is the phase pressure. Note that we have neglected gravity in the above version of the Darcy's law (gravity can be readily considered, leading to an additional term, see [**5**]). The above system can be closed by expressing the *capillary pressure* i.e, $p_c = p_w - p_o$ as a function of the saturation [**5**]. In many situations of practical interest [**5**], we are interested in the limit of zero capillary pressure. One way to formally derive the resulting equation is to assume that the capillary pressure is zero. In this case, we can sum (1.4) for both phases and obtain

$$\mathbf{v} = -\lambda_T(s) \nabla_{\mathbf{x}} p, \tag{1.5}$$

with $p = p_w = p_o$ being the pressure and $\lambda_T = \lambda_w + \lambda_o$ being the total mobility. Using (1.5), the gradient of the pressure in (1.4) can be eliminated, which yields

$$\mathbf{v}_w = \frac{\lambda_w(s)}{\lambda_T(s)} \mathbf{v}.$$

Define the fractional flow function f as

$$f(s) = \frac{\lambda_w(s)}{\lambda_T(s)} = \frac{\lambda_w(s)}{\lambda_w(s) + \lambda_o(s)}, \tag{1.6}$$

then the saturation equation (1.2) for water can be written as

$$\partial_t s + \mathrm{div}_{\mathbf{x}}(f(s)\mathbf{v}) = 0. \tag{1.7}$$

Combining the saturation equation with the incompressibility condition (1.3) and the pressure equation, we obtain the evolution equations for two phase flow in a porous medium:

$$\begin{aligned} \partial_t s + \mathrm{div}_{\mathbf{x}}(f(s)\mathbf{v}) &= 0, \\ \mathrm{div}_{\mathbf{x}}(\mathbf{v}) &= 0, \\ \mathbf{v} &= -\lambda_T(s) \nabla_{\mathbf{x}} p. \end{aligned} \tag{1.8}$$

The above equations have to be augmented by suitable initial and boundary conditions.

The phase mobility $\lambda_w : [0,1] \mapsto \mathbb{R}$ is a monotone increasing function with $\lambda_w(0) = 0$ and the phase mobility $\lambda_o : [0,1] \mapsto \mathbb{R}$ is a monotone decreasing function with $\lambda_o(1) = 0$. Furthermore, the total mobility is strictly positive, i.e, $\lambda_T \geq \lambda_* > 0$ for some λ_*. The system (1.8) is a nonlinear system of partial differential equations with the saturation equation in (1.8) a scalar hyperbolic conservation law in several space dimensions with a coefficient given by the velocity $\mathbf{v}$. The velocity can be obtained by solving an elliptic equation for the pressure p.

It is well known that solutions of hyperbolic conservation laws can develop discontinuities, even for smooth initial data, see e.g. [**73**]. The presence of these discontinuities or shock waves implies that solutions of conservation laws are sought in a weak sense and are augmented with additional admissibility criteria or *entropy conditions* in order to ensure uniqueness.

As the two phase flow equations (1.8) involve a conservation law, we need to define a suitable concept of entropy solutions for these equations and show that these solutions are well-posed. The problem of proving well-posedness of global weak solutions of the two phase flow equations (1.8) has remained open for many decades. The main challenge in showing existence is the fact that the velocity field $\mathbf{v}$ acts as a coefficient in the saturation equation. Although conservation laws with coefficients have been studied extensively in recent years, see [**1, 77, 56, 30, 3, 32**] and references therein, the state of the art results require that the coefficient is a function of bounded variation. Many attempts at showing that the velocity field $\mathbf{v}$ in (1.7) is sufficiently regular, for example is a BV function or has enough Sobolev regularity, have failed. Partial results (with strong assumptions on the velocity field or on the solution) have been obtained in [**94, 102**] and references therein.

Another approach is to consider a modified version of the two phase flow equations, recalling that the two phase flow equations (1.8) were derived under the assumption that the capillary pressure was zero. Adding small but non-zero capillary pressure leads to a viscous perturbation of the saturation equation, see [**86**]. The viscous problem has been shown to be well-posed in [**86**]. However, the fact that the coefficient of viscosity can be very small leads to difficulties in numerical approximation of these equations as the viscous scales have to resolved.

A different approach to the above two considers the more fundamental question – is the Darcy's law (1.4) correct? Many studies have focused on this question and have found that the Darcy's law may be inadequate to explain the dynamics of fluid flow in porous media, even for a single phase [**17**]. It is plausible that the problems of showing well-posedness for the full two-phase flow model can be attributed to the modeling deficiencies of the Darcy's law.

Several modifications of the Darcy's law have been proposed, see [**29**] and references therein. Of particular interest in this paper is the Brinkman modification [**17**]. It has been widely accepted in the literature that this modification explains the dynamics of flow in porous media better than the Darcy model in many situations of interest, for both one-phase as well as multi-phase flows, see [**95, 84**] and references therein. The Brinkman model for the phase velocity of each phase is given by,

$$-\mu\Delta_{\mathbf{x}}\mathbf{v}_r + \mathbf{v}_r = -\lambda_r\nabla_{\mathbf{x}}p_r, \quad r \in \{w, o\}. \tag{1.9}$$

Here, μ denotes a small scale parameter which is assumed to be identical for both phases. Note that the Brinkman approximation adds a smoothing term to Darcy's law.

Adding the phase velocity relations (1.9) for both phases w, o and neglecting capillary pressure i.e, $p_w = p_o = p$, we obtain that the total velocity $\mathbf{v} = \mathbf{v}_w + \mathbf{v}_o$ satisfies,

$$-\mu\Delta_{\mathbf{x}}\mathbf{v} + \mathbf{v} = -\lambda_T(s)\nabla_{\mathbf{x}}p.$$

Applying the divergence operator to both sides of the above equation and using incompressibility (1.3), we obtain the following elliptic equation for the pressure

$$-\mathrm{div}_{\mathbf{x}}\left(\lambda_T(s)\nabla_{\mathbf{x}}p\right) = 0.$$

Combining this equation with the conservation of mass for the aqueous phase and with the Brinkman approximation (1.9) describing the velocity of the aqueous phase and the fractional flow function defined in (1.6), we obtain the following complete system,

$$\begin{aligned} \partial_t s + \mathrm{div}_{\mathbf{x}}\left(\mathbf{v}_w\right) &= 0, \\ -\mu\Delta_{\mathbf{x}}\mathbf{v}_w + \mathbf{v}_w &= -f(s)\lambda_T(s)\nabla_{\mathbf{x}}p, \\ -\mathrm{div}_{\mathbf{x}}\left(\lambda_T(s)\nabla_{\mathbf{x}}p\right) &= 0, \end{aligned} \tag{1.10}$$

which models two-phase flow in a porous medium, where the velocity of each phase obeys the Brinkman's law (1.9). The system (1.10) is henceforth termed the Brinkman regularization of two phase flow in a porous medium. We remark that the Darcy system (1.8) can be obtained from the Brinkman regularization (1.10) by setting $\mu = 0$ and rewriting the water phase velocity in terms of the fractional flow function.

The rest of this paper is concerned with the analysis and numerical approximation of the Brinkman regularization (1.10). Our aims are threefold:

- To define a suitable notion of solutions to the Brinkman regularization (1.10) and to show that such solutions exist.
- To design an efficient numerical scheme to approximate the Brinkman regularization for two phase flows and to prove that this scheme converges when the mesh is refined.
- To compare the solutions of the Brinkman regularization with those of the standard Darcy model for two phase flow (1.8) in order to examine whether the Brinkman regularization is a suitable approximation of the Darcy's law in the regime of two phase flows.

The rest of this paper provides answers to the above questions and is organized as follows: in Section 2, equivalent forms of the Brinkman regularization are stated, a suitable notion of solutions is defined and the main existence theorem is described. Section 3 deals with the proof of existence of the Brinkman regularization. A convergent numerical scheme for approximating (1.10) is presented in Section 4. Finally, we provide further comparisons between the Darcy and Brinkman models (particularly in one space dimension) in Section 5. Some useful inequalities are collected in an Appendix.

2. Statement of problem

In this section, we consider the following Darcy-Brinkman system (1.10) augmented with initial and boundary conditions,

$$
\begin{cases}
\partial_t s + \operatorname{div}_{\mathbf{x}}(\mathbf{v}_w) = 0, & t>0,\ \mathbf{x}\in\Omega,\\
-\mu\Delta_{\mathbf{x}}\mathbf{v}_w + \mathbf{v}_w = -f(s)\lambda_T(s)\nabla_{\mathbf{x}}p, & t>0,\ \mathbf{x}\in\Omega,\\
-\operatorname{div}_{\mathbf{x}}(\lambda_T(s)\nabla_{\mathbf{x}}p) = 0, & t>0,\ \mathbf{x}\in\Omega,\\
\lambda_T(s(t,\mathbf{x}))\partial_\nu p(t,\mathbf{x}) = \pi(t,\mathbf{x}), & t>0,\ \mathbf{x}\in\partial\Omega,\\
\mathbf{v}_w(t,\mathbf{x})\cdot\nu(\mathbf{x}) = h(t,\mathbf{x}), & t>0,\ \mathbf{x}\in\partial\Omega,\\
\partial_\nu\mathbf{v}_w(t,\mathbf{x})\cdot\tau(\mathbf{x}) = 0, & t>0,\ \mathbf{x}\in\partial\Omega,\\
\int_\Omega p(t,\mathbf{x})d\mathbf{x} = 0, & t>0,\\
s(0,\mathbf{x}) = s_0(\mathbf{x}), & \mathbf{x}\in\Omega,
\end{cases}
\tag{2.1}
$$

where

- (**H.1**) Ω is an open connected subset of $\mathbb{R}^N$, $N\geq 1$, with smooth boundary $\partial\Omega$. The vector ν is the unit outer normal, and τ is any vector in the tangent plane;
- (**H.2**) f is a smooth Lipschitz bounded function, $0<\mu\leq 1$ is a constant, and h, $\pi : (0,\infty)\times\partial\Omega\to\mathbb{R}$ are smooth bounded maps;
- (**H.3**) λ_T is smooth Lipschitz bounded such that $\lambda_T(\cdot)\geq\lambda_*$ for some constant $\lambda_*>0$, and $\lambda_T f'$ and λ_T'/λ_T are bounded;
- (**H.4**) the initial datum $s_0\in H^2(\Omega)$, $0\leq s_0\leq 1$.

Note that all the above assumptions are consistent with the definitions of the phase mobilities in the Darcy's law.

Formally applying the Helmholtz operator $-\mu\Delta_{\mathbf{x}}+1$ to the first equation in (2.1) we obtain the third order problem

$$
\begin{cases}
\partial_t s - \mu\Delta_{\mathbf{x}}\partial_t s - \operatorname{div}_{\mathbf{x}}(f(s)\lambda_T(s)\nabla_{\mathbf{x}}p) = 0, & t>0,\ \mathbf{x}\in\Omega,\\
-\operatorname{div}_{\mathbf{x}}(\lambda_T(s)\nabla_{\mathbf{x}}p) = 0, & t>0,\ \mathbf{x}\in\Omega.
\end{cases}
\tag{2.2}
$$

Remark 2.1. The above equation (2.2) is very similar to a model of dynamic capillary pressure considered in [**63**], [**67**] and references therein. Thus, two independent regularization mechanisms do lead to very similar regularization terms for the transport equation of the saturation.

Since we can rewrite the first equation in the form

$$
\partial_t s - \operatorname{div}_{\mathbf{x}}(\mu\nabla_{\mathbf{x}}\partial_t s + f(s)\lambda_T(s)\nabla_{\mathbf{x}}p) = 0,
\tag{2.3}
$$

the boundary condition on $\partial_\nu\partial_t s$ reads as the flux boundary condition on (2.3). Indeed the flux in (2.3) is $-(\mu\nabla_{\mathbf{x}}\partial_t s + f(s)\lambda_T(s)\nabla_{\mathbf{x}}p)$, multiplying by the unit outer normal ν and using the fact that $\lambda_T(s)\partial_\nu p=\pi$ we have

$$
\begin{aligned}
\big(\mu\nabla_{\mathbf{x}}\partial_t s + f(s)\lambda_T(s)\nabla_{\mathbf{x}}p\big)\cdot\nu\big|_{\partial\Omega} =& \big(\mu\partial_\nu\partial_t s + f(s)\underbrace{\lambda_T(s)\partial_\nu p}_{=\pi}\big)\big|_{\partial\Omega}\\
=& \mu\partial_\nu\partial_t s + f(s)\pi.
\end{aligned}
$$

Next, we introduce the notion of weak solutions to the Brinkman system (2.1) above.

DEFINITION 2.2. We say that a triplet $(s, \mathbf{v}_w, p)$ is a weak solution of (2.1) if s, $p : [0,\infty) \times \Omega \to \mathbb{R}$, $\mathbf{v}_w : [0,\infty) \times \Omega \to \mathbb{R}^N$, and

(**D**.**1**) for every $T > 0$

$$s \in H^1(0,T;H^1(\Omega)), \quad p \in L^\infty(0,T;H^1(\Omega)), \quad \mathbf{v}_w \in L^\infty(0,T;H^2(\Omega));$$

(**D**.**2**) for every test function $\varphi \in C^\infty(\mathbb{R}^{N+1})$ with compact support, the following identity is satisfied

$$\int_0^\infty \int_\Omega (s\partial_t\varphi + \mathbf{v}_w \cdot \nabla_{\mathbf{x}}\varphi)\, d\mathbf{x}dt - \int_0^\infty \int_{\partial\Omega} h\varphi\, d\sigma dt + \int_\Omega s_0(x)\varphi(0,x)d\mathbf{x} = 0,$$
$$\int_0^\infty \int_\Omega \lambda_T(s)\nabla_{\mathbf{x}}p \cdot \nabla_{\mathbf{x}}\varphi\, d\mathbf{x}dt - \int_0^\infty \int_{\partial\Omega} \pi\varphi\, d\sigma dt = 0;$$

(**D**.**3**) for every test function $\Phi \in C^\infty(\mathbb{R} \times \Omega; \mathbb{R}^N)$ with compact support, the following identity is satisfied

$$\mu \int_0^\infty \int_\Omega \nabla_{\mathbf{x}}\mathbf{v}_w : \nabla_{\mathbf{x}}\Phi\, d\mathbf{x}dt + \int_0^\infty \int_\Omega \mathbf{v}_w \cdot \Phi\, d\mathbf{x}dt + \int_0^\infty \int_\Omega f(s)\lambda_T(s)\nabla_{\mathbf{x}}p \cdot \Phi\, d\mathbf{x}dt = 0;$$

(**D**.**4**) for almost every $t > 0$

$$\int_\Omega p(t,\mathbf{x})d\mathbf{x} = 0;$$

(**D**.**5**) for almost every $t > 0$ the boundary conditions on $\mathbf{v}_w$ are satisfied in the sense of traces.

Due to regularity assumption (**D**.**1**) and the linearity of the Helmholtz operator $\mathrm{Id} - \mu\Delta_{\mathbf{x}}$, the solutions of (2.1) solve (2.2) and vice versa. We note that the saturation is required to have some Sobolev regularity. This is in contrast to the Darcy based standard two-phase flow model where the saturation is merely required to be bounded and may contain discontinuities.

One of our main results is the following existence theorem.

THEOREM 2.3. *Assume* (**H**.**1**), (**H**.**2**), (**H**.**3**), *and* (**H**.**4**). *Then, the initial boundary value problem* (2.1) *has a solution* $(s, p, \mathbf{v}_w)$ *in the sense of Definition* 2.2.

Following [**29**], we use the following vanishing viscosity approximation of (2.1)

$$
\begin{cases}
\partial_t s_\varepsilon + \operatorname{div}_{\mathbf{x}}(\mathbf{v}_{w,\epsilon}) = \epsilon \Delta_{\mathbf{x}} s_\varepsilon, & t>0,\ \mathbf{x}\in\Omega,\\
-\mu\Delta_{\mathbf{x}}\mathbf{v}_{w,\epsilon} + \mathbf{v}_{w,\epsilon} = -f(s_\varepsilon)\lambda_T(s_\varepsilon)\nabla_{\mathbf{x}} p_\varepsilon, & t>0,\ \mathbf{x}\in\Omega,\\
-\operatorname{div}_{\mathbf{x}}(\lambda_T(s_\varepsilon)\nabla_{\mathbf{x}} p_\varepsilon) = 0, & t>0,\ \mathbf{x}\in\Omega,\\
\partial_\nu s_\varepsilon(t,\mathbf{x}) = 0, & t>0,\ \mathbf{x}\in\partial\Omega,\\
\lambda_T(s_\varepsilon(t,\mathbf{x}))\partial_\nu p_\varepsilon(t,\mathbf{x}) = \pi(t,\mathbf{x}), & t>0,\ \mathbf{x}\in\partial\Omega,\\
\mathbf{v}_{w,\epsilon}(t,\mathbf{x})\cdot\nu(\mathbf{x}) = h(t,\mathbf{x}), & t>0,\ \mathbf{x}\in\partial\Omega,\\
\partial_\nu \mathbf{v}_{w,\epsilon}(t,\mathbf{x})\cdot\tau(\mathbf{x}) = 0, & t>0,\ \mathbf{x}\in\partial\Omega,\\
\int_\Omega p_\varepsilon(t,\mathbf{x})d\mathbf{x} = 0, & t>0,\\
s_\varepsilon(0,\mathbf{x}) = s_0(\mathbf{x}), & \mathbf{x}\in\Omega,
\end{cases}
\tag{2.4}
$$

where $0<\epsilon<1$. Here, we have $N+2$ unknowns $(s_\varepsilon, \mathbf{v}_{w,\epsilon}, p_\varepsilon) : [0,\infty)\times\Omega \to \mathbb{R}\times\mathbb{R}^N\times\mathbb{R}$ and $N+2$ boundary conditions because the tangent spaces to $\partial\Omega$ have dimension $N-1$.

THEOREM 2.4. *For every $\epsilon>0$, the parabolic-elliptic boundary value problem* (2.4) *admits a unique smooth solution $(s_\varepsilon, \mathbf{v}_{w,\epsilon}, p_\varepsilon)$.*

Since the proof of this result follows a classical (tedious and long) argument we simply sketch it, see e.g. [**89, 25, 28**]. Using the regularity of the initial condition and the Contraction Mapping Principle it is possible to prove the existence of a small time $T>0$ and of a unique solution $(s_\varepsilon, \mathbf{v}_{w,\epsilon}, p_\varepsilon)$ defined in $[0,T)\times\Omega$. The a priori estimates of the next section prevent the blow-up of the solution, that is indeed defined for every time. Finally, a bootstrap argument gives the smoothness of the solution.

In the next section in order to prove Theorem 2.3 we will show the compactness of the above approximate solutions.

3. A priori estimates and proof of Theorem 2.3

This section is devoted to the proof of Theorem 2.3. We begin with some a priori estimates on the solution $(s_\varepsilon,\ \mathbf{v}_{w,\epsilon},\ p_\varepsilon)$ of (2.4).

LEMMA 3.1 (H^1 estimate on p_ε). *We have that*

$$\{p_\varepsilon\}_{\epsilon>0} \text{ is uniformly bounded in } L^\infty(0,T;H^1(\Omega)),\ T>0.$$

More precisely,

$$\|p_\varepsilon(t,\cdot)\|_{H^1(\Omega)} \le C_1 \|\pi(t,\cdot)\|_{L^2(\partial\Omega)}, \qquad t>0, \tag{3.1}$$

for some positive constant C_1 independent of μ and ϵ.

PROOF. From the third equation in (2.4), (**H.3**), and the boundary conditions on p_ε,

$$
\begin{aligned}
\lambda_* \int_\Omega |\nabla_{\mathbf{x}} p_\varepsilon|^2 d\mathbf{x} &\le \int_\Omega \lambda_T(s_\varepsilon)|\nabla_{\mathbf{x}} p_\varepsilon|^2\, d\mathbf{x}\\
&= -\int_\Omega \underbrace{\operatorname{div}_{\mathbf{x}}(\lambda_T(s_\varepsilon)\nabla_{\mathbf{x}} p_\varepsilon)}_{=0} p_\varepsilon\, d\mathbf{x} + \int_{\partial\Omega} p_\varepsilon \underbrace{\lambda_T(s_\varepsilon)\partial_\nu p_\varepsilon}_{=\pi}\, d\sigma
\end{aligned}
$$

$$= \int_{\partial\Omega} \pi p_\varepsilon \, d\sigma \leq \frac{1}{2\alpha} \int_{\partial\Omega} \pi^2 \, d\sigma + \frac{\alpha}{2} \int_{\partial\Omega} p_\varepsilon^2 \, d\sigma,$$

where $\alpha > 0$ is a constant that will be chosen later. The zero mean condition on p_ε and the Poincaré inequality give

$$\begin{aligned}\lambda_* \int_\Omega |\nabla_{\mathbf{x}} p_\varepsilon|^2 \, d\mathbf{x} &\leq \frac{1}{2\alpha} \int_{\partial\Omega} \pi^2 d\sigma + \frac{\alpha}{2} \int_{\partial\Omega} p_\varepsilon^2 \, d\sigma \\ &\leq \frac{1}{2\alpha} \int_{\partial\Omega} \pi^2 d\sigma + \frac{\alpha c}{2} \int_\Omega |\nabla_{\mathbf{x}} p_\varepsilon|^2 \, d\mathbf{x},\end{aligned}$$

where c is a constant. Choosing $\alpha = \lambda_*/c$ we get

$$\int_\Omega |\nabla_{\mathbf{x}} p_\varepsilon|^2 d\mathbf{x} \leq \frac{c}{\lambda_*^2} \int_{\partial\Omega} \pi^2 \, d\sigma.$$

The bound (3.1) follows from the zero mean condition on p_ε. □

LEMMA 3.2. *We have that*

$$\{\mathbf{v}_{w,\epsilon}\}_{\epsilon>0} \text{ is uniformly bounded in } L^\infty(0,T;H^2(\Omega)),\ T>0.$$

More precisely,

$$\|\mathbf{v}_{w,\epsilon}(t,\cdot)\|_{H^2(\Omega)} \leq \frac{C_3}{\mu} \left(\|f\|_{L^\infty(\mathbb{R})} \|\pi(t,\cdot)\|_{L^2(\partial\Omega)} + \|h(t,\cdot)\|_{L^2(\partial\Omega)} \right), \tag{3.2}$$

for each $t > 0$ and some positive constant C_3 independent of μ and ϵ.

PROOF. The claim follows directly from classical regularity results on elliptic equations [**2**, Theorem 8.2] and Lemma 3.1. □

LEMMA 3.3. *We have that*

$$\begin{aligned}&\{s_\varepsilon\}_{\epsilon>0} \text{ is uniformly bounded in } H^1(0,T;H^1(\Omega)),,\ T>0,\\ &\{\sqrt{\epsilon}\Delta_{\mathbf{x}} s_\varepsilon\}_{\epsilon>0} \text{ is uniformly bounded in } L^2((0,T)\times\Omega)),\ T>0.\end{aligned}$$

In particular

$$\begin{aligned}\|s_\varepsilon(t,\cdot)\|^2_{H^1(\Omega)} + \epsilon e^{2t} \int_0^t e^{-2\tau} \|\Delta_{\mathbf{x}} s_\varepsilon(\tau,\cdot)\|^2_{L^2(\Omega)} \, d\tau \leq \|s_0\|^2_{H^1(\Omega)} e^{2t} \\ + \frac{C_4}{\mu^2} e^{2t} \int_0^t e^{-2\tau} \left(\|f\|^2_{L^\infty(\mathbb{R})} \|\pi(\tau,\cdot)\|^2_{L^2(\partial\Omega)} + \|h(\tau,\cdot)\|^2_{L^2(\partial\Omega)} \right) d\tau,\end{aligned} \tag{3.3}$$

$$\begin{aligned}\|\partial_t s_\varepsilon\|^2_{L^2(0,T;H^1(\Omega))} \leq \|\nabla_{\mathbf{x}} s_0\|^2_{H^1(\Omega)} \\ + \frac{C_4}{\mu^2} \left(\|f\|^2_{L^\infty(\mathbb{R})} \|\pi\|^2_{L^2((0,T)\times\partial\Omega)} + \|h\|^2_{L^2((0,T)\times\partial\Omega)} \right),\end{aligned} \tag{3.4}$$

for each $t > 0$ and some positive constant C_4 independent of μ and ϵ.

PROOF. Multiplying the first equation in (2.4) by $s_\varepsilon - \Delta_{\mathbf{x}} s_\varepsilon$ and integrating over Ω we get

$$\begin{aligned}\int_\Omega \partial_t s_\varepsilon (s_\varepsilon - \Delta_{\mathbf{x}} s_\varepsilon) d\mathbf{x} =& \epsilon \int_\Omega s_\varepsilon \Delta_{\mathbf{x}} s_\varepsilon \, d\mathbf{x} - \epsilon \int_\Omega |\Delta_{\mathbf{x}} s_\varepsilon|^2 d\mathbf{x} \\ &- \int_\Omega \operatorname{div}_{\mathbf{x}} (\mathbf{v}_{w,\epsilon})\, s_\varepsilon d\mathbf{x} + \int_\Omega \operatorname{div}_{\mathbf{x}} (\mathbf{v}_{w,\epsilon})\, \Delta_{\mathbf{x}} s_\varepsilon \, d\mathbf{x}.\end{aligned} \tag{3.5}$$

We have

$$\begin{aligned}\int_\Omega \partial_t s_\varepsilon (s_\varepsilon - \Delta_{\mathbf{x}} s_\varepsilon) d\mathbf{x} =& \frac{d}{dt} \int_\Omega \frac{s_\varepsilon^2 + |\nabla_{\mathbf{x}} s_\varepsilon|^2}{2} \, d\mathbf{x}, \\ \epsilon \int_\Omega s_\varepsilon \Delta_{\mathbf{x}} s_\varepsilon \, d\mathbf{x} \le& \frac{\epsilon}{2} \int_\Omega s_\varepsilon^2 \, d\mathbf{x} + \frac{\epsilon}{2} \int_\Omega |\Delta_{\mathbf{x}} s_\varepsilon|^2 \, d\mathbf{x} \\ \le& \frac{1}{2} \int_\Omega s_\varepsilon^2 \, d\mathbf{x} + \frac{\epsilon}{2} \int_\Omega |\Delta_{\mathbf{x}} s_\varepsilon|^2 d\mathbf{x}, \qquad \text{(recall that } \epsilon < 1), \\ -\int_\Omega \operatorname{div}_{\mathbf{x}} (\mathbf{v}_{w,\epsilon})\, s_\varepsilon \, d\mathbf{x} \le& \frac{1}{2} \int_\Omega s_\varepsilon^2 \, d\mathbf{x} + c_1 \left\| \mathbf{v}_{w,\epsilon}(t,\cdot) \right\|^2_{H^1(\Omega)}, \\ \int_\Omega \operatorname{div}_{\mathbf{x}} (\mathbf{v}_{w,\epsilon})\, \Delta_{\mathbf{x}} s_\varepsilon \, d\mathbf{x} \le& \frac{1}{2} \int_\Omega |\nabla_{\mathbf{x}} s_\varepsilon|^2 \, d\mathbf{x} + c_2 \left\| \mathbf{v}_{w,\epsilon}(t,\cdot) \right\|^2_{H^2(\Omega)},\end{aligned}$$

for some constants c_1, $c_2 > 0$. Therefore, using (3.2), from (3.5) we deduce

$$\begin{aligned}\frac{d}{dt} \int_\Omega \frac{s_\varepsilon^2 + |\nabla_{\mathbf{x}} s_\varepsilon|^2}{2} \, d\mathbf{x} \le 2 \int_\Omega \frac{s_\varepsilon^2 + |\nabla_{\mathbf{x}} s_\varepsilon|^2}{2} \, d\mathbf{x} - \frac{\epsilon}{2} \int_\Omega |\Delta_{\mathbf{x}} s_\varepsilon|^2 \, d\mathbf{x} \\ + \frac{c_3}{\mu^2} \left(\|f\|^2_{L^\infty(\mathbb{R})} \left\| \pi(t,\cdot) \right\|^2_{L^2(\partial\Omega)} + \left\| h(t,\cdot) \right\|^2_{L^2(\partial\Omega)} \right),\end{aligned}$$

for some constant $c_3 > 0$. Clearly, (3.3) follows from the Gronwall's inequality.

Multiplying the first equation in (2.4) by $\partial_t s_\varepsilon - \partial_t \Delta_{\mathbf{x}} s_\varepsilon$ and integrating over Ω we get

$$\begin{aligned}\int_\Omega \partial_t s_\varepsilon (\partial_t s_\varepsilon - \partial_t \Delta_{\mathbf{x}} s_\varepsilon) d\mathbf{x} =& \epsilon \int_\Omega \partial_t s_\varepsilon \Delta_{\mathbf{x}} s_\varepsilon d\mathbf{x} - \epsilon \int_\Omega \Delta_{\mathbf{x}} s_\varepsilon \partial_t \Delta_{\mathbf{x}} s_\varepsilon d\mathbf{x} \\ &- \int_\Omega \operatorname{div}_{\mathbf{x}} (\mathbf{v}_{w,\epsilon})\, \partial_t s_\varepsilon d\mathbf{x} + \int_\Omega \operatorname{div}_{\mathbf{x}} (\mathbf{v}_{w,\epsilon})\, \partial_t \Delta_{\mathbf{x}} s_\varepsilon d\mathbf{x}.\end{aligned} \tag{3.6}$$

Since

$$\partial_t \partial_\nu s_\varepsilon(t,\mathbf{x}) = 0, \quad t > 0,\, \mathbf{x} \in \partial\Omega,$$

we have

$$\begin{aligned}\int_\Omega \partial_t s_\varepsilon (\partial_t s_\varepsilon - \partial_t \Delta_{\mathbf{x}} s_\varepsilon) d\mathbf{x} &= \int_\Omega \left((\partial_t s_\varepsilon)^2 + |\partial_t \nabla_{\mathbf{x}} s_\varepsilon|^2 \right) d\mathbf{x}, \\ \epsilon \int_\Omega \partial_t s_\varepsilon \Delta_{\mathbf{x}} s_\varepsilon \, d\mathbf{x} &= -\frac{\epsilon}{2} \frac{d}{dt} \int_\Omega |\nabla_{\mathbf{x}} s_\varepsilon|^2 d\mathbf{x}, \\ -\epsilon \int_\Omega \partial_t \Delta_{\mathbf{x}} s_\varepsilon \Delta_{\mathbf{x}} s_\varepsilon \, d\mathbf{x} &= -\frac{\epsilon}{2} \frac{d}{dt} \int_\Omega |\Delta_{\mathbf{x}} s_\varepsilon|^2 \, d\mathbf{x},\end{aligned}$$

$$-\int_\Omega \operatorname{div}_{\mathbf{x}}(\mathbf{v}_{w,\epsilon})\,\partial_t s_\varepsilon d\mathbf{x} \le \frac{1}{2}\int_\Omega (\partial_t s_\varepsilon)^2\,d\mathbf{x} + c_4\,\|\mathbf{v}_{w,\epsilon}(t,\cdot)\|^2_{H^1(\Omega)},$$
$$\int_\Omega \operatorname{div}_{\mathbf{x}}(\mathbf{v}_{w,\epsilon})\,\partial_t\Delta_{\mathbf{x}} s_\varepsilon\,d\mathbf{x} \le \frac{1}{2}\int_\Omega |\partial_t\nabla_{\mathbf{x}} s_\varepsilon|^2 d\mathbf{x} + c_5\,\|\mathbf{v}_{w,\epsilon}(t,\cdot)\|^2_{H^2(\Omega)},$$

for some constants c_4, $c_5 > 0$. Therefore integrating over $(0,\tau)$, (3.6) implies

$$\int_0^\tau\int_\Omega \left((\partial_t s_\varepsilon)^2 + |\partial_t\nabla_{\mathbf{x}} s_\varepsilon|^2\right) dtd\mathbf{x} + \epsilon\int_\Omega \left(|\nabla_{\mathbf{x}} s_\varepsilon(\tau,x)|^2 + |\Delta_{\mathbf{x}} s_\varepsilon(\tau,x)|^2\right) d\mathbf{x}$$
$$\le \int_\Omega \left(|\nabla_{\mathbf{x}} s_0|^2 + |\Delta_{\mathbf{x}} s_0|^2\right) d\mathbf{x} + \frac{c_6}{\mu^2}\left(\|f\|^2_{L^\infty(\mathbb{R})}\,\|\pi\|^2_{L^2((0,\tau)\times\partial\Omega)} + \|h\|^2_{L^2((0,\tau)\times\partial\Omega)}\right),$$

for some constant $c_6 > 0$. This proves (3.4). □

LEMMA 3.4. *We have that*

$$\{\Delta_{\mathbf{x}} p_\varepsilon\}_{\epsilon>0} \text{ is uniformly bounded in } L^\infty(0,T;L^1(\Omega)),\ T>0.$$

More precisely,

$$\|\Delta_{\mathbf{x}} p_\varepsilon(t,\cdot)\|_{L^1(\Omega)} \le C_5\left(\|\pi(t,\cdot)\|^2_{L^2(\partial\Omega)} + \|h(t,\cdot)\|^2_{L^2(\partial\Omega)}\right),$$

for each $t > 0$, where C_5 is a positive constant independent of μ and ϵ.

PROOF. From (2.4)

$$\Delta_{\mathbf{x}} p_\varepsilon = -\frac{\lambda_T'(s_\varepsilon)}{\lambda_T(s_\varepsilon)}\nabla_{\mathbf{x}} p_\varepsilon\cdot\nabla_{\mathbf{x}} s_\varepsilon,$$

therefore, thanks to (**H**.**2**) and (**H**.**3**),

$$\|\Delta_{\mathbf{x}} p_\varepsilon(t,\cdot)\|_{L^1(\Omega)} \le \frac{1}{2}\left\|\frac{\lambda_T'}{\lambda_T}\right\|_{L^\infty(\mathbb{R})}\left(\|\nabla_{\mathbf{x}} p_\varepsilon(t,\cdot)\|^2_{L^2(\Omega)} + \|\nabla_{\mathbf{x}} s_\varepsilon(t,\cdot)\|^2_{L^2(\Omega)}\right).$$

The claim follows from Lemmas 3.1 and 3.3 and using that $H^1(0,T;H^1(\Omega)) \subset L^\infty(0,T;H^1(\Omega))$ for $T > 0$. □

PROOF OF THEOREM 2.3. Thanks to Lemmas 3.1, 3.2, 3.3, there exists a sequence $\{\varepsilon_k\}_{k\in\mathbb{N}}$, $\varepsilon_k \to 0$, and three functions

$$s,\ p : (0,\infty)\times\Omega \longrightarrow \mathbb{R}, \qquad \mathbf{v}_w : (0,\infty)\times\Omega \longrightarrow \mathbb{R}^N, \tag{3.7}$$

such that, for every $T > 0$,

$$s \in H^1(0,T;H^1(\Omega)), \quad p \in L^\infty(0,T;H^1(\Omega)), \quad \mathbf{v}_w \in L^\infty(0,T;H^2(\Omega)),$$

and

$$\begin{aligned} s_{\varepsilon_k} &\rightharpoonup s, && \text{weakly in } H^1(0,T;H^1(\Omega)),\ T>0,\\ p_{\varepsilon_k} &\rightharpoonup p, && \text{weakly in } L^\ell(0,T;H^1(\Omega)),\ 1\le\ell<\infty,\ T>0,\\ \mathbf{v}_{w,\varepsilon_k} &\rightharpoonup \mathbf{v}_w, && \text{weakly in } L^\ell(0,T;H^2(\Omega)),\ 1\le\ell<\infty,\ T>0. \end{aligned} \tag{3.8}$$

In particular, we have that

$$
\begin{aligned}
s_{\varepsilon_k} &\to s, && \text{strongly in } L^2((0,T)\times\Omega),\ T>0,\\
s_{\varepsilon_k} &\to s, && \text{a.e. in } (0,\infty)\times\Omega,\\
\nabla_{\mathbf{x}} p_{\varepsilon_k} &\rightharpoonup \nabla_{\mathbf{x}} p, && \text{weakly in } L^2((0,T)\times\Omega),\ T>0.
\end{aligned}
\tag{3.9}
$$

Let $\varphi \in C^\infty(\mathbb{R}^{N+1})$ be a test function with compact support. We have

$$
\int_0^\infty\int_\Omega (s_{\varepsilon_k}\partial_t\varphi + \mathbf{v}_{w,\varepsilon_k}\cdot\nabla_{\mathbf{x}}\varphi)\,d\mathbf{x}dt - \int_0^\infty\int_{\partial\Omega} h\varphi\,d\sigma dt + \int_\Omega s_0(x)\varphi(0,x)\,d\mathbf{x} = \varepsilon_k\int_0^\infty\nabla_{\mathbf{x}} s_{\varepsilon_k}\nabla_{\mathbf{x}}\varphi\,dtd\mathbf{x},
$$

and

$$
\int_0^\infty\int_\Omega \lambda_T(s_{\varepsilon_k})\nabla_{\mathbf{x}} p_{\varepsilon_k}\cdot\nabla_{\mathbf{x}}\varphi\,d\mathbf{x}dt - \int_0^\infty\int_{\partial\Omega}\pi\varphi\,d\sigma dt = 0.
$$

Therefore, the dominated convergence theorem, (3.9), and the boundedness of λ_T imply (**D**.**2**). Moreover, for every test function $\Phi \in C^\infty(\mathbb{R}\times\Omega;\mathbb{R}^N)$ with compact support, we have

$$
\mu\int_0^\infty\int_\Omega \nabla_{\mathbf{x}}\mathbf{v}_{w,\varepsilon_k} : \nabla_{\mathbf{x}}\Phi\,d\mathbf{x}dt + \int_0^\infty\int_\Omega \mathbf{v}_{w,\varepsilon_k}\cdot\Phi\,d\mathbf{x}dt + \int_0^\infty\int_\Omega f(s_{\varepsilon_k})\lambda_T(s_{\varepsilon_k})\nabla_{\mathbf{x}} p\cdot\Phi\,d\mathbf{x}dt = 0.
$$

Therefore, the dominated convergence theorem, (3.9), and the boundedness of f and λ_T imply (**D**.**3**). Finally, for every test function $\phi \in C^\infty(\mathbb{R})$ with compact support, we have

$$
\int_0^\infty\int_\Omega \phi(t)p_{\varepsilon_k}(t,\mathbf{x})dtd\mathbf{x} = 0, \qquad t>0.
$$

Therefore, (3.8) implies (**D**.**4**).

We conclude by noting that (**D**.**1**) holds thanks to Lemmas 3.1, 3.2 and 3.3.

□

Thus, we have shown that weak solutions of the Brinkman regularization of two-phase flows in a porous medium (1.10) exist. The question of uniqueness is still open.

Remark 3.5. It must be emphasized that many of the estimates derived in the proof of the existence theorem 2.3 are μ dependent and blow up as the regularization parameter $\mu \to 0$. In particular, the estimate (3.2) on the phase velocity is μ-dependent as are the estimates on the saturation (3.3), (3.4). Thus, in the limit $\mu \to 0$, which corresponds to the classical Darcy's law, we do not expect that the velocity field and the saturation are as regular as in the case of the Brinkman approximation. As an example, it is well known that the saturation contains discontinuities in the form of shocks for the classical two-phase flow problem, something which is inconsistent with the H^1 estimate in (3.3), (3.4). Hence, we have been unable to obtain any convergence results for the Brinkman system (1.10) to the classical two-phase Darcy system (1.8) as $\mu \to 0$.

REMARK 3.6. Here, we have focused on the case of two-phase flows. A Brinkman regularization of multi-phase flows can be obtained analogously to the derivation of the Brinkman two phase flow model in the introduction. This system for m ($m \geq 3$) phases reads as

$$\begin{cases} \partial_t s_1 + \operatorname{div}_{\mathbf{x}}(\mathbf{v}_{w,1}) = 0, & t > 0,\ \mathbf{x} \in \Omega, \\ \cdots\cdots \\ \partial_t s_m + \operatorname{div}_{\mathbf{x}}(\mathbf{v}_{w,m}) = 0, & t > 0,\ \mathbf{x} \in \Omega, \\ -\mu\Delta_{\mathbf{x}}\mathbf{v}_{w,1} + \mathbf{v}_{w,1} = -\lambda_1(s_1)\nabla_{\mathbf{x}} p, & t > 0,\ \mathbf{x} \in \Omega, \\ \cdots\cdots \\ -\mu\Delta_{\mathbf{x}}\mathbf{v}_{w,m} + \mathbf{v}_{w,m} = -\lambda_m(s_m)\nabla_{\mathbf{x}} p, & t > 0,\ \mathbf{x} \in \Omega, \\ -\operatorname{div}_{\mathbf{x}}(\lambda_T(s_1, .., s_m)\nabla_{\mathbf{x}} p) = 0, & t > 0,\ \mathbf{x} \in \Omega, \end{cases}$$

augmented with suitable initial and boundary conditions. As in Definition 2.2, we can analogously define a suitable notion of weak solutions and prove existence of solutions by following the approximation procedure presented in Section 2 and proving analogous estimates like those in the proof of Theorem 2.3.

4. A convergent numerical scheme for the Brinkman regularization

In this section, we will present an efficient numerical scheme to approximate the Brinkman regularization for two-phase flow (1.10). For simplicity, we consider the unit square in two space dimensions, i.e, $\Omega = [0,1]^2 \subset \mathbb{R}^2$. As many interesting benchmark tests include a source in the pressure equation (to model injection of water), we consider the following modification of the Brinkman regularization (1.10),

$$\begin{cases} \partial_t s + \operatorname{div}_{\mathbf{x}}(\mathbf{v}_w) = 0, & t > 0,\ \mathbf{x} \in \Omega, \\ -\mu\Delta_{\mathbf{x}}\mathbf{v}_w + \mathbf{v}_w = -f(s)\lambda_T(s)\nabla_{\mathbf{x}} p, & t > 0,\ \mathbf{x} \in \Omega, \\ -\operatorname{div}_{\mathbf{x}}(\lambda_T(s)\nabla_{\mathbf{x}} p) = q, & t > 0,\ \mathbf{x} \in \Omega, \\ \lambda_T(s(t,\mathbf{x}))\partial_\nu p(t,\mathbf{x}) = \pi(t,\mathbf{x}), & t > 0,\ \mathbf{x} \in \partial\Omega, \\ \mathbf{v}_w(t,\mathbf{x}) \cdot \nu(\mathbf{x}) = h(t,\mathbf{x}), & t > 0,\ \mathbf{x} \in \partial\Omega, \\ \partial_\nu \mathbf{v}_w(t,\mathbf{x}) \cdot \tau(\mathbf{x}) = 0, & t > 0,\ \mathbf{x} \in \partial\Omega, \\ \int_\Omega p(t,\mathbf{x}) d\mathbf{x} = 0, & t > 0, \\ s(0,\mathbf{x}) = s_0(\mathbf{x}), & \mathbf{x} \in \Omega. \end{cases} \tag{4.1}$$

Here, $q \in L^\infty(0,T;L^2(\Omega))$ denotes a source function. Note that the existence result Theorem 2.3 can be readily extended to this case. We will assume that $\lambda_T \in L^\infty(\mathbb{R})$, i.e. $\lambda_T(x) \leq \lambda^*$, $x \in \mathbb{R}$, for this section.

For the sake of definiteness, let $\mathbf{v}_w = (u, v)$. We use the mixed Dirichlet/Neumann boundary conditions

$$\begin{aligned} &u(0,y) = u(1,y) = 0,\ \partial_y u(x,0) = \partial_y u(x,1) = 0, \\ &v(x,0) = v(x,1) = 0,\ \partial_x v(0,y) = \partial_x v(1,y) = 0, \end{aligned} \tag{4.2}$$

for $\mathbf{v}_w$, which is often used in applications. We discretize the computational domain $[0,1]^2$ by a Cartesian mesh with gridpoints $x_i = (i-1/2)\Delta x$, $y_j = (j-1/2)\Delta x$, $i,j = 1,\dots,N$, $\Delta x = 1/N$. Let p^n_{ij}, $\mathbf{v}^n_{ij}$, and s^n_{ij} denote the approximation to p, $\mathbf{v}_w$ and s respectively, evaluated at (x_i, y_j, t_n), where $t_n = n\Delta t$. Furthermore, we use the notation

$$D^x_\pm \kappa^n_{ij} = \pm\frac{1}{\Delta x}\left(\kappa^n_{i\pm1,j} - \kappa^n_{ij}\right), \quad D^y_\pm \kappa^n_{ij} = \pm\frac{1}{\Delta x}\left(\kappa^n_{i,j\pm1} - \kappa^n_{ij}\right),$$

for any grid function κ^n_{ij}.

For the saturation, similarly to the viscous approximation (2.4), we use boundary conditions

$$\begin{aligned} &s^n_{0,j} = s^n_{1,j}, \quad s^n_{N+1,j} = s^n_{N,j}, \quad j = 1,\dots,N,\\ &s^n_{i,0} = s^n_{i,1}, \quad s^n_{i,N+1} = s^n_{i,N}, \quad i = 1,\dots,N. \end{aligned} \tag{4.3}$$

The scheme for p^n_{ij} reads

$$-D^x_+\left(t^n_{i-1/2,j} D^x_- p^n_{ij}\right) - D^y_+\left(t^n_{i,j-1/2} D^y_- p^n_{ij}\right) = q^n_{ij}, \quad i,\, j = 1,\dots,N, \tag{4.4}$$

where

$$t^n_{i+1/2,j} = \frac{\lambda_T(s^n_{ij}) + \lambda_T(s^n_{i+1,j})}{2} \quad \text{and} \quad t^n_{i,j+1/2} = \frac{\lambda_T(s^n_{ij}) + \lambda_T(s^n_{i,j+1})}{2},$$

with the boundary values

$$\begin{aligned} &t^n_{i,1/2} D^y_- p^n_{i,1} = \pi^n_{i,1/2}, \quad t^n_{i,N+1/2} D^y_+ p^n_{i,N} = \pi^n_{i,N+1/2}, \quad i = 1,\dots,N,\\ &t^n_{1/2,j} D^x_- p^n_{1,j} = \pi^n_{1/2,j}, \quad t^n_{N+1/2,j} D^x_+ p^n_{N,j} = \pi^n_{N+1/2,j}, \quad j = 1,\dots,N, \end{aligned} \tag{4.5}$$

where

$$\pi^n_{i,1/2} = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \pi(t^n; x, 0)\, dx, \quad \pi^n_{i,N+1/2} = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} \pi(t^n; x, 1)\, dx, \quad i = 1,\dots,N,$$

$$\pi^n_{1/2,j} = \frac{1}{\Delta y}\int_{y_{j-1/2}}^{y_{j+1/2}} \pi(t^n; 0, y)\, dy, \quad \pi^n_{N+1/2,j} = \frac{1}{\Delta y}\int_{y_{i-1/2}}^{y_{i+1/2}} \pi(t^n; 1, y)\, dy, \quad j = 1,\dots,N,$$

as well as the constraint

$$\overline{p} := \Delta x \Delta y \sum_{i,j=1}^{N} p^n_{ij} \equiv 0, \tag{4.6}$$

and

$$q^n_{ij} = q(x_i, y_j), \quad i,j = 1,\dots,N,\ n \geq 0.$$

To define the scheme for $\mathbf{v}^n_{ij}$, we first define

$$f^n_{i+1/2,j} = \frac{f(s^n_{ij}) + f(s^n_{i+1,j})}{2} \quad \text{and} \quad f^n_{i,j+1/2} = \frac{f(s^n_{ij}) + f(s^n_{i,j+1})}{2}.$$

Then the scheme for $u^n_{i+1/2,j}$ reads

$$-\mu\left(D^x_+ D^x_- + D^y_+ D^y_-\right) u^n_{i+1/2,j} + u^n_{i+1/2,j} = -f^n_{i+1/2,j} t^n_{i+1/2,j} D^x_+ p^n_{ij}, \tag{4.7}$$

for $i = 1, \ldots, N-1$, $j = 1, \ldots, N$ with boundary values

$$u^n_{1/2,j} = u^n_{N+1/2,j} = 0, \ j = 1, \ldots, N, \text{ and}$$
$$u^n_{i+1/2,0} = u^n_{i+1/2,1}, \ u^n_{i+1/2,N+1} = u^n_{i+1/2,N}, \quad i = 1, \ldots, N-1,$$

which is a discrete version of the boundary conditions (4.2). Similarly, the scheme for $v^n_{i,j+1/2}$ reads

$$-\mu \left(D^x_+ D^x_- + D^y_+ D^y_-\right) v^n_{i,j+1/2} + v^n_{i,j+1/2} = -f^n_{i,j+1/2} t^n_{i,j+1/2} D^y_+ p^n_{ij}, \tag{4.8}$$

for $i = 1, \ldots, N$, $j = 1, \ldots, N-1$ with boundary values

$$v^n_{i,1/2} = v^n_{i,N+1/2} = 0, \ i = 1, \ldots, N, \text{ and}$$
$$v^n_{0,j+1/2} = v^n_{1,j+1/2}, \ v^n_{N+1,j+1/2} = v^n_{N,j+1/2}, \quad j = 1, \ldots, N-1.$$

Finally we update s^n_{ij} by

$$s^{n+1}_{ij} = \frac{1}{4}\left(s^n_{i+1,j} + s^n_{i-1,j} + s^n_{i,j+1} + s^n_{i,j-1}\right) - \Delta t \left(D^x_- u^n_{i+1/2,j} + D^y_- v^n_{i,j+1/2}\right), \tag{4.9}$$

for $n \geq 0$ and $i, j = 1, \ldots, N$, with the initial values $s^0_{ij} = s_0(x_i, y_j)$ and boundary conditions (4.3).

4.1. Convergence of the scheme in 2D. We will show that the approximate solutions generated by the finite difference scheme (4.4) – (4.9) converge to a weak solution of (4.1) for a fixed μ. To do so, we mimic the estimates of Lemmas 3.1 – 3.3 in the discrete setting.

From the discrete values s^n_{ij}, $i, j = 0, \ldots, N$, $n \geq 0$, we define the piecewise constant interpolant

$$s_\Delta(t; x, y) = s^n_{ij}, \quad (t; x, y) \in [t_n, t_{n+1}) \times [x_{i-1/2}, x_{i+1/2}) \times [y_{j-1/2}, y_{j+1/2}),$$
$$n \geq 0, \ i, j = 0, \ldots, N+1,$$

where we have denoted $x_{i+1/2} := i\Delta x$ and similarly $y_{j+1/2} = j\Delta y$. In a similar way, we define p_Δ, u_Δ and v_Δ to be the piecewise constant interpolations,

$$p_\Delta(t; x, y) = p^n_{ij}, \qquad (t; x, y) \in [t_n, t_{n+1}) \times [x_{i-1/2}, x_{i+1/2}) \times [y_{j-1/2}, y_{j+1/2}),$$
$$u_\Delta(t; x, y) = u^n_{i+1/2,j}, \qquad (t; x, y) \in [t_n, t_{n+1}) \times [x_i, x_{i+1}) \times [y_{j-1/2}, y_{j+1/2}),$$
$$v_\Delta(t; x, y) = v^n_{i,j+1/2}, \qquad (t; x, y) \in [t_n, t_{n+1}) \times [x_{i-1/2}, x_{i+1/2}) \times [y_j, y_{j+1}),$$

for $n \geq 0$, and $i, j = 0, \ldots, N+1$. We define them to be zero outside $(-\Delta x, 1+\Delta x) \times (-\Delta y, 1+\Delta y)$. We extend the difference operators $D^{x,y}_\pm$ to the interpolations in the obvious way,

$$D^x_\pm \sigma_\Delta(x, y) = \pm \frac{\sigma_\Delta(x \pm \Delta x, y) - \sigma_\Delta(x, y)}{\Delta x} \quad D^y_\pm \sigma_\Delta(x, y) = \pm \frac{\sigma_\Delta(x, y \pm \Delta y) - \sigma_\Delta(x, y)}{\Delta y}.$$

Moreover, we define the following discrete versions of the $H^1(\Omega)$- and $H^2(\Omega)$-norms,

$$
\begin{aligned}
|\sigma_\Delta|^2_{H^1_\Delta} &:= \Delta x \Delta y \sum_{i,j=1}^{N} \left(|D^x_- \sigma_{ij}|^2 + |D^y_- \sigma_{ij}|^2 \right), \\
\|\sigma_\Delta\|^2_{H^1_\Delta} &:= \|\sigma_\Delta\|^2_{L^2([0,1]^2)} + |\sigma_\Delta|^2_{H^1_\Delta}, \\
|\sigma_\Delta|^2_{H^2_\Delta} &:= \Delta x \Delta y \sum_{i,j=1}^{N} \left(|D^x_- D^x_+ \sigma_{ij}|^2 + |D^y_- D^y_+ \sigma_{ij}|^2 + 2|D^x_- D^y_- \sigma_{ij}|^2 \right), \\
\|\sigma_\Delta\|^2_{H^2_\Delta} &:= \|\sigma_\Delta\|^2_{H^1_\Delta} + |\sigma_\Delta|^2_{H^2_\Delta},
\end{aligned}
\tag{4.10}
$$

(replace i by $i+1/2$ for u_Δ and j by $j+1/2$ for v_Δ) and the discrete L^2-'trace'-norm

$$
\|\sigma_\Delta\|^2_{L^2(\partial_\Delta)} := \Delta x \sum_{i=1}^{N} \left((\sigma_{i0})^2 + (\sigma_{i,N+1})^2 \right) + \Delta y \sum_{j=1}^{N} \left((\sigma_{0j})^2 + (\sigma_{N+1,j})^2 \right). \tag{4.11}
$$

Now, we will show the following estimates on the approximate solutions:

LEMMA 4.1. *Let $\Delta = (\Delta x, \Delta y, \Delta t)$, $\Delta x, \Delta y, \Delta t > 0$ and $q \in L^\infty(0,T;L^2(\Omega))$. We have $p_\Delta \in L^\infty(0,T;H^1_\Delta)$ for any $T > 0$ with*

$$
\|p_\Delta(t;\cdot)\|^2_{H^1_\Delta} \le \frac{4}{(\lambda_*)^2} \left(\|q_\Delta(t;\cdot)\|^2_{L^2(\Omega)} + \left(10 + \frac{(\lambda_*)^2}{20} \Delta x \right) \|\pi(t;\cdot)\|^2_{L^2(\partial\Omega)} \right), \tag{4.12}
$$

PROOF. Using summation by parts, we have

$$
\begin{aligned}
\sum_{i=1}^{N+1} \sum_{j=1}^{N} t^n_{i-1/2,j} |D^x_- p^n_{ij}|^2 &= - \sum_{i,j=1}^{N} D^x_+ (t^n_{i-1/2,j} D^x_- p^n_{ij})\, p^n_{ij} \\
&\quad + \frac{1}{\Delta x} \sum_{j=1}^{N} \left(t^n_{N+1/2,j} p^n_{N+1,j} D^x_- p^n_{N+1,j} - t^n_{1/2,j} p_{0j} D^x_- p^n_{1j} \right), \\
\sum_{j=1}^{N+1} \sum_{i=1}^{N} t^n_{i,j-1/2} |D^y_- p^n_{ij}|^2 &= - \sum_{i,j=1}^{N} D^y_+ (t^n_{i,j-1/2} D^y_- p^n_{ij})\, p^n_{ij} \\
&\quad + \frac{1}{\Delta y} \sum_{i=1}^{N} \left(t^n_{i,N+1/2} p^n_{i,N+1} D^y_- p^n_{i,N+1} - t^n_{i,1/2} p_{i0} D^y_- p^n_{i,1} \right).
\end{aligned}
$$

Thus multiplying (4.4) by p_{ij} and summing over the indices $i, j = 1, \ldots, N$, then using the boundary conditions (4.5), we obtain

$$
\sum_{i=1}^{N+1} \sum_{j=1}^{N} t^n_{i-1/2,j} |D^x_- p^n_{ij}|^2 + \sum_{j=1}^{N+1} \sum_{i=1}^{N} t^n_{i,j-1/2} |D^y_- p^n_{ij}|^2 = \sum_{i,j=1}^{N} q^n_{ij} p^n_{ij}
$$

$$+\frac{1}{\Delta x}\sum_{j=1}^{N}\left(p^n_{N+1,j}\pi^n_{N+1/2,j}+p_{0j}\pi^n_{1/2,j}\right)+\frac{1}{\Delta y}\sum_{i=1}^{N}\left(p^n_{i,N+1}\pi^n_{i,N+1/2}+p_{i0}\pi^n_{i,1/2}\right).$$

Since $t^n_{i-1/2,j},\, t^n_{i,j-1/2}\geq\lambda_*$ by assumption (**H.3**), and $(\alpha\, a^2+b^2/\alpha)/2\geq ab$ for $a,b\in\mathbb{R}$, $\alpha>0$, this yields

$$\begin{aligned}\sum_{i,j=1}^{N}\left(|D^x_- p^n_{ij}|^2+|D^y_- p^n_{ij}|^2\right)\leq\frac{1}{2\lambda_*}\Bigg(&\alpha_1\sum_{i,j=1}^{N}(p^n_{ij})^2+\frac{1}{\alpha_1}\sum_{i,j=1}^{N}(q^n_{ij})^2\\&+\frac{1}{\Delta x}\left\{\alpha_2\sum_{j=1}^{N}\left(\left(p^n_{N+1,j}\right)^2+\left(p^n_{0j}\right)^2\right)+\frac{1}{\alpha_2}\sum_{j=1}^{N}\left(\left(\pi^n_{N+1/2,j}\right)^2+\left(\pi^n_{1/2,j}\right)^2\right)\right\}\\&+\frac{1}{\Delta y}\left\{\alpha_2\sum_{i=1}^{N}\left(\left(p^n_{i,N+1}\right)^2+\left(p^n_{i0}\right)^2\right)+\frac{1}{\alpha_2}\sum_{i=1}^{N}\left(\left(\pi^n_{i,N+1/2}\right)^2+\left(\pi^n_{i,1/2}\right)^2\right)\right\}\Bigg),\end{aligned}$$

and hence

$$|p_\Delta|^2_{H^1_\Delta}\leq\frac{1}{2\lambda_*}\left(\alpha_1\|p_\Delta\|^2_{L^2(\Omega)}+\frac{1}{\alpha_1}\|q_\Delta\|^2_{L^2(\Omega)}+\alpha_2\|p_\Delta\|^2_{L^2(\partial_\Delta)}+\frac{1}{\alpha_2}\|\pi(t;\cdot)\|^2_{L^2(\partial\Omega)}\right).$$

Using the 'discrete trace inequality', (7.1), this implies

$$\begin{aligned}|p_\Delta|^2_{H^1_\Delta}\leq\frac{1}{2\lambda_*}\Bigg(&\alpha_1\|p_\Delta\|^2_{L^2(\Omega)}+\frac{1}{\alpha_1}\|q_\Delta\|^2_{L^2(\Omega)}+2\,\alpha_2\,|p_\Delta|^2_{H^1_\Delta}+8\,\alpha_2\|p_\Delta\|^2_{L^2(\Omega)}\\&+2\alpha_2\Delta y\Delta x\left\{\sum_{j=1}^{N}\left(\pi^n_{N+1/2,j}\right)^2+\sum_{i=1}^{N}\left(\pi^n_{i,N+1/2}\right)^2\right\}+\frac{1}{\alpha_2}\|\pi(t;\cdot)\|^2_{L^2(\partial\Omega)}\Bigg).\end{aligned}$$

Then we apply the discrete Poincaré inequality (7.4) with (4.6) and obtain

$$\left(1-\frac{\alpha_1+10\alpha_2}{2\lambda_*}\right)|p_\Delta|^2_{H^1_\Delta}\leq\frac{1}{2\lambda_*}\left(\frac{1}{\alpha_1}\|q_\Delta(t;\cdot)\|^2_{L^2(\Omega)}+\left(\frac{1}{\alpha_2}+2\alpha_2\Delta x\right)\|\pi(t;\cdot)\|^2_{L^2(\partial\Omega)}\right),$$

and hence, choosing $\alpha_1=\lambda_*/2$ and $\alpha_2=\lambda_*/20$

$$|p_\Delta|^2_{H^1_\Delta}\leq\frac{2}{(\lambda_*)^2}\left(\|q_\Delta(t;\cdot)\|^2_{L^2(\Omega)}+\left(10+\frac{(\lambda_*)^2}{20}\Delta x\right)\|\pi(t;\cdot)\|^2_{L^2(\partial\Omega)}\right),$$

and (4.12) follows. □

LEMMA 4.2. *Let $\Delta=(\Delta x,\Delta y,\Delta t)$, $\Delta x,\Delta y,\Delta t>0$, $\mu>0$ and $q\in L^\infty(0,T;L^2(\Omega))$. Assume furthermore that $f\lambda_T$ is bounded. Then $u_\Delta,v_\Delta\in L^\infty(0,T;H^2_\Delta)$ for $T>0$ with*

$$\mu^2|u_\Delta(t;\cdot)|^2_{H^2_\Delta}+\mu|u_\Delta(t;\cdot)|^2_{H^1_\Delta}+\|u_\Delta(t;\cdot)\|^2_{L^2(\Omega)}\leq C\,\|f\lambda_T\|^2_{L^\infty}\|D^x_+p_\Delta(t;\cdot)\|^2_{L^2(\Omega)},\tag{4.13a}$$

$$\mu^2|v_\Delta(t;\cdot)|^2_{H^2_\Delta}+\mu|v_\Delta(t;\cdot)|^2_{H^1_\Delta}+\|v_\Delta(t;\cdot)\|^2_{L^2(\Omega)}\leq C\,\|f\lambda_T\|^2_{L^\infty}\|D^y_+p_\Delta(t;\cdot)\|^2_{L^2(\Omega)},\tag{4.13b}$$

where $C>0$ is a scaling factor, not depending on the other quantities.

PROOF. We take the square of equation (4.7), sum it over the indices i and j and use the summation by parts identity

$$\sum_{i=1}^{N-1}\sum_{j=1}^{N} u^n_{i+1/2,j}(D^x_+D^x_- + D^y_+D^y_-)u^n_{i+1/2,j} = -\sum_{i,j=1}^{N}\left(|D^x_- u^n_{i+1/2,j}|^2 + |D^y_- u^n_{i+1/2,j}|^2\right)$$

to obtain

(4.14)

$$\sum_{i,j=1}^{N}\left(\mu^2|(D^x_+D^x_- + D^y_+D^y_-)u^n_{i+1/2,j}|^2 + 2\mu(|D^x_- u^n_{i+1/2,j}|^2 + |D^y_- u^n_{i+1/2,j}|^2) + |u^n_{i+1/2,j}|^2\right) = \sum_{i,j=1}^{N}(f^n_{i+1/2,j})^2(t^n_{i+1/2,j})^2|D^x_+ p^n_{ij}|^2.$$

Using summation by parts twice for the first term on the left hand side of (4.14) and the boundary conditions gives

$$\sum_{i,j=1}^{N}|(D^x_+D^x_- + D^y_+D^y_-)u^n_{i+1/2,j}|^2 = \sum_{i,j=1}^{N}\left(|D^x_+D^x_- u^n_{i+1/2,j}|^2 + |D^y_+D^y_- u^n_{i+1/2,j}|^2 + 2|D^x_-D^y_- u^n_{i+1/2,j}|^2\right),$$

which implies (4.13a). In the same way, we can show (4.13b). Since $p_\Delta \in L^\infty(0,T;H^1_\Delta)$ by Lemma 4.1, we obtain $u_\Delta, v_\Delta \in L^\infty(0,T;H^2_\Delta)$. □

Before the next lemma, we need some additional notation: We define

$$D^t\sigma_\Delta(t;x,y) := \frac{1}{\Delta t}\left(\sigma_\Delta(t+\Delta t;x,y) - \sigma_\Delta(t;x,y)\right),$$

and a discrete version of the $W^{1,\infty}(0,T;L^2(\Omega))$-norm,

$$\|s_\Delta\|^2_{W^{1,\infty}_{\Delta t}(0,T;L^2(\Omega))} := \sup_{t\in[0,T]}\left\{\|s_\Delta(t;\cdot)\|^2_{L^2(\Omega)} + \|D^t s_\Delta(t;\cdot)\|^2_{L^2(\Omega)}\right\}.$$

Now it is easy to show that $s_\Delta \in W^{1,\infty}_{\Delta t}(0,T;L^2(\Omega)) \cap L^\infty(0,T;H^1_\Delta)$:

LEMMA 4.3. *Let $\Delta = (\Delta x, \Delta y, \Delta t)$, $\Delta x, \Delta y, \Delta t > 0$ with $\Delta x/\Delta t, \Delta y/\Delta t \le K$, where $0 < K < \infty$, and let $\mu > 0$. Moreover assume that $q \in L^\infty(0,T;L^2(\Omega))$, $s_0 \in H^1(\Omega)$ and that $f\lambda_T$ is bounded. Define*

$$\Gamma_{f,\pi,q} = \|f\|_{L^\infty}\left(\|\pi\|_{L^\infty(0,T;L^2(\partial\Omega))} + \|\lambda_T\|_{L^\infty}\|q\|_{L^\infty(0,T;L^2(\Omega))}\right).$$

Then $s_\Delta \in W^{1,\infty}_{\Delta t}(0,T;L^2(\Omega)) \cap L^\infty(0,T;H^1_\Delta)$ for $T > 0$ with

$$\|s_\Delta(t;\cdot)\|_{L^2(\Omega)} \le \|s_0\|_{L^2(\Omega)} + \frac{C\,t}{\sqrt{\mu}}\Gamma_{f,\pi,q} \tag{4.15a}$$

$$|s_\Delta(t;\cdot)|_{H^1_\Delta} \le |s_0|_{H^1(\Omega)} + \frac{C\,t}{\mu}\,\Gamma_{f,\pi,q} \tag{4.15b}$$

$$\|D^t s_\Delta(t;\cdot)\|_{L^2} \le C\left(\sqrt{K}+1\right)\left(|s_0|_{H^1} + \frac{1+t}{\mu}\,\Gamma_{f,\pi,q}\right) \tag{4.15c}$$

for $0 \le t \le T$ and where $C > 0$ is a constant.

Proof. We take the square of equation (4.9), sum over the indices i, j and use the triangle inequality to obtain

$$\left(\sum_{i,j=1}^{N} |s_{ij}^{n+1}|^2\right)^{1/2} \le \left(\sum_{i,j=1}^{N} |s_{ij}^{n}|^2\right)^{1/2} + \Delta t\left(\sum_{i,j=1}^{N} |D_-^x u_{i+1/2,j}^n + D_-^y v_{i,j+1/2}^n|^2\right)^{1/2},$$

which implies

$$\|s_\Delta(t_{n+1};\cdot)\|_{L^2(\Omega)} \le \|s_\Delta(t_n;\cdot)\|_{L^2(\Omega)} + \Delta t\big(\|D_-^x u_\Delta(t_n;\cdot)\|_{L^2(\Omega)} + \|D_-^y v_\Delta(t_n;\cdot)\|_{L^2(\Omega)}\big).$$

Iterating over n, this yields

$$\|s_\Delta(t_n;\cdot)\|_{L^2(\Omega)} \le \|s_0\|_{L^2(\Omega)} + t_n\big(\sup_{n\ge0}|u_\Delta(t_n;\cdot)|_{H^1_\Delta} + \sup_{n\ge0}|v_\Delta(t_n;\cdot)|_{H^1_\Delta}\big).$$

Using Lemma 4.1 and 4.2, we obtain

$$\begin{aligned}\|s_\Delta(t;\cdot)\|_{L^2(\Omega)} &\le \|s_0\|_{L^2(\Omega)} + \frac{C\,t}{\sqrt{\mu}}\|f\lambda_T\|_{L^\infty}\sup_{n\ge0}|p_\Delta(t_n;\cdot)|_{H^1_\Delta}\\ &\le \|s_0\|_{L^2(\Omega)} + \frac{C\,t}{\sqrt{\mu}}\|f\|_{L^\infty}\left(\|\lambda_T\|_{L^\infty}\|q\|_{L^\infty(0,T;L^2(\Omega))} + \|\pi\|_{L^\infty(0,T;L^2(\partial\Omega))}\right),\end{aligned}$$

where we have used (4.12) for the second inequality. In order to show that $s_\Delta(t;\cdot)$ is in H^1_Δ, i.e. that the gradient of $s_\Delta(t;\cdot)$ is in $L^2(\Omega)$, we apply the linear operators D_+^x, D_+^y to the evolution equation for s_{ij}^n, (4.9),

$$\begin{aligned}D_+^x s_{ij}^{n+1} = \frac{1}{4}\left(D_+^x s_{i+1,j}^n + D_+^x s_{i-1,j}^n + D_+^x s_{i,j+1}^n + D_+^x s_{i,j-1}^n\right)\\ - \Delta t\left(D_+^x D_-^x u_{i+1/2,j}^n + D_+^x D_-^y v_{i,j+1/2}^n\right),\end{aligned}$$

(and similarly for D_+^y), then take the square of the above equation, sum over the indices i, j and use again triangle inequality, to obtain

$$\left(\sum_{i,j=1}^{N} |D_+^x s_{ij}^{n+1}|^2\right)^{1/2} \le \left(\sum_{i,j=1}^{N} |D_+^x s_{ij}^{n}|^2\right)^{1/2} + \Delta t\left(\sum_{i,j=1}^{N} |D_+^x D_-^x u_{i+1/2,j}^n + D_+^x D_-^y u_{i,j+1/2}^n|^2\right)^{1/2},$$

which implies after iteration over n,

$$\left(\sum_{i,j=1}^{N} |D_+^x s_{ij}^{n}|^2\right)^{1/2} \le \left(\sum_{i,j=1}^{N} |D_+^x s_{ij}^{0}|^2\right)^{1/2} + t_n \sup_{n\ge0}\left(\sum_{i,j=1}^{N} |D_+^x D_-^x u_{i+1/2,j}^n + D_+^x D_-^y u_{i,j+1/2}^n|^2\right)^{1/2},$$

A similar estimate holds for the differences $D_+^y s_{ij}^n$ and hence, using Lemmas 4.1 and 4.2, we obtain

$$|s_\Delta(t;\cdot)|_{H^1_\Delta} \le |s_0|_{H^1_\Delta} + \frac{C\,t}{\mu}\|f\lambda_T\|_{L^\infty(\mathbb{R})}\left(\|q\|_{L^\infty(0,T;L^2(\Omega))} + \|\pi\|_{L^\infty(0,T;L^2(\partial\Omega))}\right).$$

To obtain an estimate on $D^t s_\Delta$, we rewrite the evolution equation for s_{ij}^n as

$$\frac{s_{ij}^{n+1} - s_{ij}^n}{\Delta t} = \frac{1}{4\Delta t}\left(\Delta x^2 D_+^x D_-^x s_{ij}^n + \Delta y^2 D_+^y D_-^y s_{ij}^n\right) - \left(D_-^x u_{i+1/2,j}^n + D_-^y v_{i,j+1/2}^n\right), \tag{4.16}$$

We notice that

$$\Delta x|D_+^x D_-^x s_{ij}^n| \le |D_+^x s_{ij}^n| + |D_-^x s_{ij}^n|,$$
$$\Delta y|D_+^y D_-^y s_{ij}^n| \le |D_+^y s_{ij}^n| + |D_-^y s_{ij}^n|,$$

which implies after taking the square of equation (4.16) and summing over i, j

$$\sum_{i,j=1}^N |D^t s_{ij}^n|^2 \le K\sum_{i,j=1}^N \left(|D_+^x s_{ij}^n|^2 + |D_+^y s_{ij}^n|^2\right) + 2\sum_{i,j=1}^N \left(|D_-^x u_{i+1/2,j}^n + D_-^y u_{i,j+1/2}^n|^2\right).$$

Thus

$$\begin{aligned}
&\|D^t s_\Delta(t;\cdot)\|_{L^2(\Omega)} \\
&\quad\le C\left(\sqrt{K}|s_\Delta|_{L^\infty(0,T;H^1_\Delta)} + |u_\Delta|_{L^\infty(0,T;H^1_\Delta)} + |v_\Delta|_{L^\infty(0,T;H^1_\Delta)}\right) \\
&\quad\le C\left(\sqrt{K}\|\nabla_{\mathbf{x}} s_0\|_{L^2(\Omega)} + \frac{\sqrt{K}t+1}{\lambda_*\mu}\|f\lambda_T\|_{L^\infty}\left(\|q\|_{L^\infty(0,T;L^2(\Omega))} + \|\pi\|_{L^\infty(0,T;L^2(\partial\Omega))}\right)\right),
\end{aligned}$$

where we have used (4.15b) and Lemmas 4.1 and 4.2 for the second inequality. □

Now we are ready to prove the main convergence theorem for the finite difference scheme.

THEOREM 4.4. *Fix $\mu > 0$ and assume $q \in L^\infty(0,T;L^2(\Omega))$, $s_0 \in H^1(\Omega)$ and $f, \lambda_T \in L^\infty(\mathbb{R})$. Furthermore, let $\Delta = (\Delta x, \Delta y, \Delta t) > 0$ such that $\Delta x/\Delta t, \Delta y/\Delta t \le K < \infty$. Then a subsequence of $\{p_\Delta\}_{\Delta>0}$, $\{u_\Delta\}_{\Delta>0}$, $\{v_\Delta\}_{\Delta>0}$, $\{s_\Delta\}_{\Delta>0}$, converges to a weak solution $(p, \mathbf{v}_w, s)$ of (4.1) as $\Delta \to 0$, and*

$$s \in L^\infty(0,T;H^1(\Omega)) \cap W^{1,\infty}(0,T;L^2(\Omega)) \quad p \in L^\infty(0,T;H^1(\Omega)), \quad \mathbf{v}_w \in L^\infty(0,T;H^2(\Omega)).$$

PROOF. Due to Lemmas 4.1, 4.2 and 4.3, we have,

$$\sup_{t\in[0,T]} \|p_\Delta\|_{H^1_\Delta} \le C$$
$$\sup_{t\in[0,T]} \left(\|u_\Delta\|_{H^2_\Delta} + \|v_\Delta\|_{H^2_\Delta}\right) \le C$$
$$\sup_{t\in[0,T]} \left(\|s_\Delta\|_{H^1_\Delta} + \|D^t s_\Delta\|_{L^2(\Omega)}\right) \le C,$$

for a constant C independent of Δ. Then it follows from Ladyzenskaja's theorems of interpolation of finite difference approximations [**88**, Lemma 3.1, 3.2, Theorem 3.2], that

$$\begin{aligned} &s_\Delta \rightharpoonup s, && \text{weakly in } H^1([0,T]\times\Omega),\ T>0,\\ &s_\Delta \to s, && \text{strongly in } L^2((0,T)\times\Omega),\ T>0, \text{ and}\\ &s_\Delta \to s, && \text{a.e. in } (0,T)\times\Omega, \end{aligned} \tag{4.17}$$

with the limit $s \in L^\infty(0,T;H^1(\Omega)) \cap W^{1,\infty}(0,T;L^2(\Omega))$. Lemmas 3.1., 3.2 from [**88**] can easily be generalized to other L^p-spaces so that we also have

$$\begin{aligned} p_\Delta &\rightharpoonup p, && \text{weakly in } L^\ell(0,T;L^2(\Omega)),\ T>0,\ 1\le\ell<\infty,\\ (D^x_- p_\Delta, D^y_- p_\Delta)^T &\rightharpoonup \nabla_{\mathbf{x}} p, && \text{weakly in } L^\ell(0,T;L^2(\Omega)),\ T>0,\ 1\le\ell<\infty, \end{aligned} \tag{4.18}$$

with $p \in L^\infty(0,T;H^1(\Omega))$. Similarly, thanks to [**88**, Theorem 4.2],

$$\begin{aligned} (u_\Delta, v_\Delta) &\rightharpoonup (u,v), && \text{weakly in } L^\ell(0,T;L^2(\Omega)),\ T>0,\ 1\le\ell<\infty,\\ (D^x_\pm u_\Delta, D^y_\pm u_\Delta)^T &\rightharpoonup \nabla_{\mathbf{x}} u, && \text{weakly in } L^\ell(0,T;L^2(\Omega)),\ T>0,\ 1\le\ell<\infty,\\ (D^x_\pm v_\Delta, D^y_\pm v_\Delta)^T &\rightharpoonup \nabla_{\mathbf{x}} v, && \text{weakly in } L^\ell(0,T;L^2(\Omega)),\ T>0,\ 1\le\ell<\infty,\\ D^{z_1}_\pm D^{z_2}_\pm u_\Delta, &\rightharpoonup \partial_{z_1}\partial_{z_2} u, && \text{weakly in } L^\ell(0,T;L^2(\Omega)),\ T>0,\ 1\le\ell<\infty,\\ D^{z_1}_\pm D^{z_2}_\pm v_\Delta, &\rightharpoonup \partial_{z_1}\partial_{z_2} v, && \text{weakly in } L^\ell(0,T;L^2(\Omega)),\ T>0,\ 1\le\ell<\infty, \end{aligned} \tag{4.19}$$

($z_i \in \{x,y\}$, $i=1,2$) with $\mathbf{v}_w := (u,v) \in L^\infty(0,T;H^2(\Omega))$ We denote

$$\Lambda_\Delta = \begin{pmatrix} \lambda^x_\Delta & 0 \\ 0 & \lambda^y_\Delta \end{pmatrix}, \quad F_\Delta = \begin{pmatrix} f^x_\Delta & 0 \\ 0 & f^y_\Delta \end{pmatrix},$$

where

$$\begin{aligned} \lambda^x_\Delta(t;x,y) &= t^n_{i+1/2,j}, && (t;x,y) \in [t_n,t_{n+1})\times[x_i,x_{i+1})\times[y_{j-1/2},y_{j+1/2}),\\ \lambda^y_\Delta(t;x,y) &= t^n_{i,j+1/2}, && (t;x,y) \in [t_n,t_{n+1})\times[x_{i-1/2},x_{i+1/2})\times[y_j,y_{j+1}),\\ f^x_\Delta(t;x,y) &= f^n_{i+1/2,j}, && (t;x,y) \in [t_n,t_{n+1})\times[x_i,x_{i+1})\times[y_{j-1/2},y_{j+1/2}),\\ f^y_\Delta(t;x,y) &= f^n_{i,j+1/2}, && (t;x,y) \in [t_n,t_{n+1})\times[x_{i-1/2},x_{i+1/2})\times[y_j,y_{j+1}), \end{aligned}$$

Then, thanks to (4.17),

$$\begin{aligned} \lambda^{x,y}_\Delta &\to \lambda_T(s), && \text{a.e. in } (0,T)\times\Omega,\\ f^{x,y}_\Delta &\to f(s), && \text{a.e. in } (0,T)\times\Omega. \end{aligned} \tag{4.20}$$

From the definition of the scheme, (4.4), (4.7), (4.8) and (4.9) and the definitions of p_Δ, u_Δ etc.

(4.21)

$$\begin{aligned} &- D^x_+(\lambda^x_\Delta(x-\Delta x/2,y)D^x_- p_\Delta(x,y)) - D^y_+(\lambda^y_\Delta(x,y-\Delta y/2)D^y_- p_\Delta(x,y)) = q_\Delta(x,y),\\ &- \mu(D^x_+D^x_- + D^y_+D^y_-)u_\Delta(x,y) + u_\Delta(x,y) = -f^x_\Delta(x,y)\lambda^x_\Delta(x,y)D^x_+ p_\Delta(x-\Delta x/2,y),\\ &- \mu(D^x_+D^x_- + D^y_+D^y_-)v_\Delta(x,y) + v_\Delta(x,y) = -f^y_\Delta(x,y)\lambda^y_\Delta(x,y)D^y_+ p_\Delta(x,y-\Delta y/2), \end{aligned}$$

and

$$D^t s_\Delta(t;x,y) + D^x_- u_\Delta(t;x+\Delta x/2,y) + D^y_- v_\Delta(t;x,y+\Delta y/2) = \frac{1}{4\Delta t}(\Delta x^2 D^x_- D^x_+ + \Delta y^2 D^y_- D^y_+) s_\Delta(t;x,y). \tag{4.22}$$

To simplify, we introduce the notations,

$$\nabla_h g = \begin{pmatrix} D^x_- g \\ D^y_- g \end{pmatrix}, \quad \nabla_h \begin{pmatrix} a^1 \\ a^2 \end{pmatrix} = \begin{pmatrix} D^x_- a^1 & D^y_- a^1 \\ D^x_- a^2 & D^y_- a^2 \end{pmatrix}$$

for functions $g, a^{1,2}$ defined on $[0,T]\times\Omega$, and

$$\widetilde{\Lambda}_\Delta(t;x,y) = \begin{pmatrix} \widetilde{\lambda}^x_\Delta(t;x,y) & 0 \\ 0 & \widetilde{\lambda}^y_\Delta(t;x,y) \end{pmatrix},$$

where $\widetilde{\lambda}^x_\Delta(t;x,y) = \lambda^x_\Delta(t;x-\Delta x/2,y)$, $\widetilde{\lambda}^y_\Delta(t;x,y) = \lambda^y_\Delta(t;x,y-\Delta y/2)$ and in the same way, we define $\widetilde{F}_\Delta$. Moreover, we let

$$\pi_\Delta(t;x,y) = \begin{cases} \pi^n_{1/2,j}, & (t;x,y)\in[t_n,t_{n+1})\times[-\Delta x,0]\times[y_{j-1/2},y_{j+1/2}), \\ \pi^n_{N+1/2,j}, & (t;x,y)\in[t_n,t_{n+1})\times[1,1+\Delta x]\times[y_{j-1/2},y_{j+1/2}), \\ \pi^n_{i,1/2}, & (t;x,y)\in[t_n,t_{n+1})\times[x_{i-1/2},x_{i+1/2}]\times[-\Delta y,0), \\ \pi^n_{i,N+1/2}, & (t;x,y)\in[t_n,t_{n+1})\times[x_{i-1/2},x_{i+1/2})\times[1,1+\Delta y). \end{cases}$$

for $i=1,\dots,N$, $j=1,\dots,N$ and $n\geq 0$. Then multiplying the equations (4.21) and (4.22) by test functions $\varphi\in C^\infty_0([0,T)\times\mathbb{R}^2)$ and $\Phi\in C^\infty([0,T)\times\mathbb{R}^2;\mathbb{R}^2)$ with compact support, integrating over $[0,T]\times\Omega$ and relabeling the integration variables, we have

$$\begin{aligned} &-\int_0^T\int_\Omega \big(s_\Delta D^t_-\varphi + (u_\Delta(\cdot;x-\Delta x/2,\cdot), v_\Delta(\cdot;\cdot,y-\Delta y/2))^T\cdot\nabla_h\varphi \\ &\qquad -\frac{1}{4\Delta t}\nabla_h s_\Delta\cdot(\Delta x^2 D^x_-\varphi, \Delta y^2 D^y_-\varphi)^T\big)\,d\mathbf{x}\,dt - \int_\Omega s_\Delta(0;\cdot)\varphi(0;\cdot)\,d\mathbf{x} = 0, \\ &\frac{1}{\Delta x}\int_0^T\int_0^1\left\{\int_{-\Delta x}^0 \pi_\Delta\varphi dx - \int_1^{1+\Delta x}\pi_\Delta\varphi(\cdot;x-\Delta x,\cdot)dx\right\}dy\,dt \\ &\quad + \frac{1}{\Delta y}\int_0^T\int_0^1\left\{\int_{-\Delta y}^0 \pi_\Delta\varphi dy - \int_1^{1+\Delta y}\pi_\Delta\varphi(\cdot;\cdot,y-\Delta y)dy\right\}dx\,dt \\ &\qquad + \int_0^T\int_\Omega\left(\left(\widetilde{\Lambda}_\Delta\nabla_h p_\Delta\right)\cdot\nabla_h\varphi - q_\Delta\varphi\right)d\mathbf{x}\,dt = 0; \end{aligned} \tag{4.23}$$

where $D^t_-\varphi(t;\cdot) = (\varphi(t;\cdot)-\varphi(t-\Delta t;\cdot))/\Delta t$, and

$$\mu\int_0^T\int_\Omega \nabla_h(u_\Delta(\cdot;x-\Delta x/2,\cdot), v_\Delta(\cdot;\cdot,y-\Delta y/2))^T : \nabla_h\Phi\,d\mathbf{x}\,dt + \int_0^T\int_\Omega (u_\Delta(\cdot;x-\Delta x/2,\cdot), v_\Delta(\cdot;\cdot,y-\Delta y/2))^T\cdot\Phi\,d\mathbf{x}\,dt \tag{4.24}$$

$$= -\int_0^T \int_\Omega \left(\widetilde{F}_\Delta \widetilde{\Lambda}_\Delta \nabla_h p_\Delta \right) \cdot \Phi \, d\mathbf{x} \, dt;$$

Using (4.17), the smoothness of φ and the CFL-condition, we find, as $\Delta \to 0$,

$$\begin{aligned} -\int_0^T \int_\Omega s_\Delta D_-^t \varphi \, d\mathbf{x} \, dt &\to -\int_0^T \int_\Omega s \varphi_t d\mathbf{x} \, dt, \\ -\int_\Omega s_\Delta(0;\cdot)\varphi(0;\cdot) \, d\mathbf{x} &\to -\int_\Omega s_0 \, \varphi(0;\cdot) \, d\mathbf{x}, \\ -\frac{1}{4\Delta t} \int_0^T \int_\Omega \nabla_h s_\Delta \cdot (\Delta x^2 D_-^x \varphi, \Delta y^2 D_-^y \varphi)^T \, d\mathbf{x} \, dt &\to 0. \end{aligned} \tag{4.25}$$

Furthermore, using (4.19), (4.18), (4.20) and again the smoothness of φ and Φ, we have letting $\Delta \to 0$,

(4.26)

$$\begin{aligned} -\int_0^T \int_\Omega (u_\Delta(\cdot; x - \Delta x/2, \cdot), v_\Delta(\cdot;\cdot, y - \Delta y/2))^T \cdot \nabla_h \varphi \, d\mathbf{x} \, dt &\to -\int_0^T \int_\Omega \mathbf{v}_w \cdot \nabla_{\mathbf{x}} \varphi \, d\mathbf{x} \, dt, \\ -\int_0^T \int_\Omega \left(\widetilde{F}_\Delta \widetilde{\Lambda}_\Delta \nabla_h p_\Delta \right) \cdot \Phi \, d\mathbf{x} \, dt &\to -\int_0^T \int_\Omega f(s)\lambda(s) \nabla_{\mathbf{x}} p \cdot \Phi \, d\mathbf{x} \, dt, \\ \int_0^T \int_\Omega \left(\left(\widetilde{\Lambda}_\Delta \nabla_h p_\Delta \right) \cdot \nabla_h \varphi - q_\Delta \varphi \right) d\mathbf{x} \, dt &\to \int_0^T \int_\Omega \left(\lambda(s) \nabla_{\mathbf{x}} p \cdot \nabla_{\mathbf{x}} \varphi - q \, \varphi \right) d\mathbf{x} \, dt, \end{aligned}$$

and

$$\begin{aligned} \mu \int_0^T \int_\Omega \nabla_h (u_\Delta(\cdot; x - \Delta x/2, \cdot), v_\Delta(\cdot;\cdot, y - \Delta y/2))^T : \nabla_h \Phi \, d\mathbf{x} \, dt \\ + \int_0^T \int_\Omega (u_\Delta(\cdot; x - \Delta x/2, \cdot), v_\Delta(\cdot;\cdot, y - \Delta y/2))^T \cdot \Phi \, d\mathbf{x} \, dt \\ \to \int_0^T \int_\Omega \left(\mu \nabla_{\mathbf{x}} \mathbf{v}_w : \nabla_{\mathbf{x}} \Phi + \mathbf{v}_w \cdot \Phi \right) d\mathbf{x} dt. \end{aligned} \tag{4.27}$$

Furthermore, by (4.20), and since φ is smooth and $\pi \in L^2([0,T] \times \partial\Omega)$, the boundary terms converge as $\Delta \to 0$:

$$\begin{aligned} \frac{1}{\Delta x} \int_0^T \int_0^1 \left\{ \int_{-\Delta x}^0 \pi_\Delta \varphi dx - \int_1^{1+\Delta x} \pi_\Delta \varphi(\cdot; x - \Delta x, \cdot) dx \right\} dy \, dt \\ + \frac{1}{\Delta y} \int_0^T \int_0^1 \left\{ \int_{-\Delta y}^0 \pi_\Delta \varphi dy - \int_1^{1+\Delta y} \pi_\Delta \varphi(\cdot;\cdot, y - \Delta y) dy \right\} dx \, dt \\ \to \int_0^T \int_{\partial\Omega} \pi \varphi \, d\sigma \, dt. \end{aligned} \tag{4.28}$$

Summing up, (4.23), (4.24), (4.25), (4.26), (4.27) and (4.28) imply that $(s, p, \mathbf{v}_w)$ is a weak solution of the equations. □

4.2. Numerical experiments. We will now show through numerical experiments that the finite difference scheme (4.4) – (4.9) is effective in computing approximate solutions of the Brinkman regularization of the two-phase flow problem (4.1). We consider the well-known *quarter five spot* problem that models water flooding in an oil reservoir. To this end, we consider

$$q(\mathbf{x}) = \begin{cases} 4/(\pi r^2), & |\mathbf{x}| \leq r, \\ -4/(\pi r^2), & |\mathbf{x} - (1,1)| \leq r, \\ 0, & \text{otherwise}, \end{cases}$$

where $r = 0.02$. This models the injection of water at $(0,0)$ and the production of oil at $(1,1)$. The initial water saturation was given by

$$s_0(\mathbf{x}) = \begin{cases} 1, & |\mathbf{x}| \leq r, \\ \exp\left(-150(|\mathbf{x}| - r)^2\right), & |\mathbf{x}| > r. \end{cases}$$

Furthermore, the boundary values of the saturation are given by,

$$s^n_{0,j} = s^n_{1,j},\ s^n_{N+1,j} = s^n_{N,j},\ s^n_{i,0} = s^n_{1,0},\ \text{and}\ s^n_{i,N+1} = s^n_{i,N},$$

as well as

$$s^{n+1}_{ij} = 1 \ \text{ if } |(x_i, y_j)| \leq r. \tag{4.29}$$

4.2.1. *Convergence tests for a fixed μ.* We consider the Brinkman regularization with a fixed $\mu = 0.005$ and compute the approximate saturation with the numerical scheme (4.4) – (4.9), on a sequence of meshes ranging from 100×100 to 800×800 mesh points. The results of the water saturation at $t = 1$ are shown in Figure 1. The results show that the saturation is computed in a robust manner and converges. The limit seems to consist of a series of waves emanating from the injection in the lower left corner.

4.2.2. *Effect of the vanishing regularization parameter μ.* The regularization parameter μ serves to indicate the deviation of the regularized problem from the classical Darcy two-phase flow problem (1.8). Formally, we can recover the classical two-phase flow problem from the regularized Brinkman approximation by letting $\mu \to 0$. On the other hand, we were unable to rigorously establish whether such a limit exists and whether it is also a weak solution of the classical two-phase flow problem (1.8), see Remark 3.1. Hence, we will investigate this issue numerically by considering the quarter-five spot problem as in the previous experiment for different values of the regularization parameter μ. We present the water saturation at time $t = 0.65$ on a 2000×2000 grid, computed for four different values, $\mu = \{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$. The results are shown in Figure 2. Two features in the results stand out. First, the solutions become very oscillatory (at least near the injection corner) as μ is reduced and the saturation is no longer in the physically relevant $s \in [0,1]$ range. Second, the solutions consist of a moving front between $s = 0$ and $s = 1$, followed by a train of oscillatory waves. The above results are clearly consistent with the theory. The stability estimates on the regularized saturation and velocity are μ dependent (see Remark 3.1) and blow up as $\mu \to 0$. Furthermore, the convergence results for the scheme hold for any fixed non-zero μ and the stability estimates for the scheme break down as $\mu \to 0$. This break

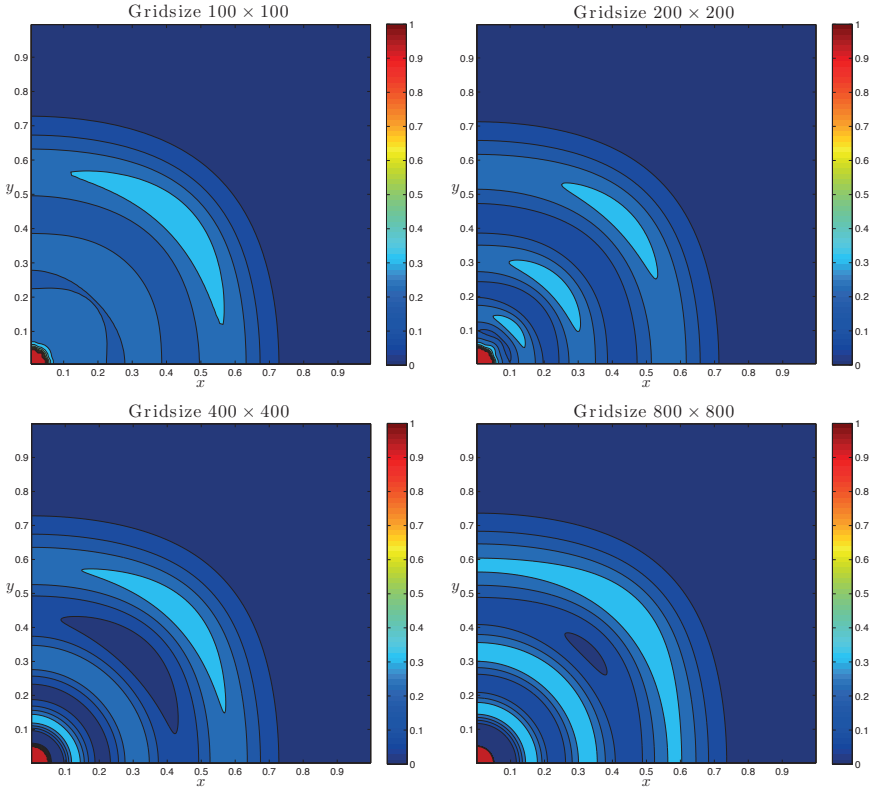

FIGURE 1. Water saturation at time $t = 1$, computed with the finite difference scheme (4.4) – (4.9) on a sequence of nested meshes with fixed regularization parameter $\mu = 0.005$.

down of the estimates is perhaps reflected in the high-frequency oscillations that arise in the numerical solution as $\mu \to 0$. This clearly indicates that the zero regularization limit may not be well-posed and the solutions of the Brinkman regularization may not converge to a weak solution of the Darcy based two-phase problem (1.8) as $\mu \to 0$.

5. Analysis in one space dimension

In order to further investigate whether the zero μ limit of the regularized Brinkman equation (1.10) converges to the Darcy two-phase flow equations (1.8), we consider the highly simplified case of one space dimension, i.e, $\Omega \subset \mathbb{R}$. In this case, the pressure equation can be solved, and the solution normalized so that $\lambda_T(s)p_x = 1$. This gives the

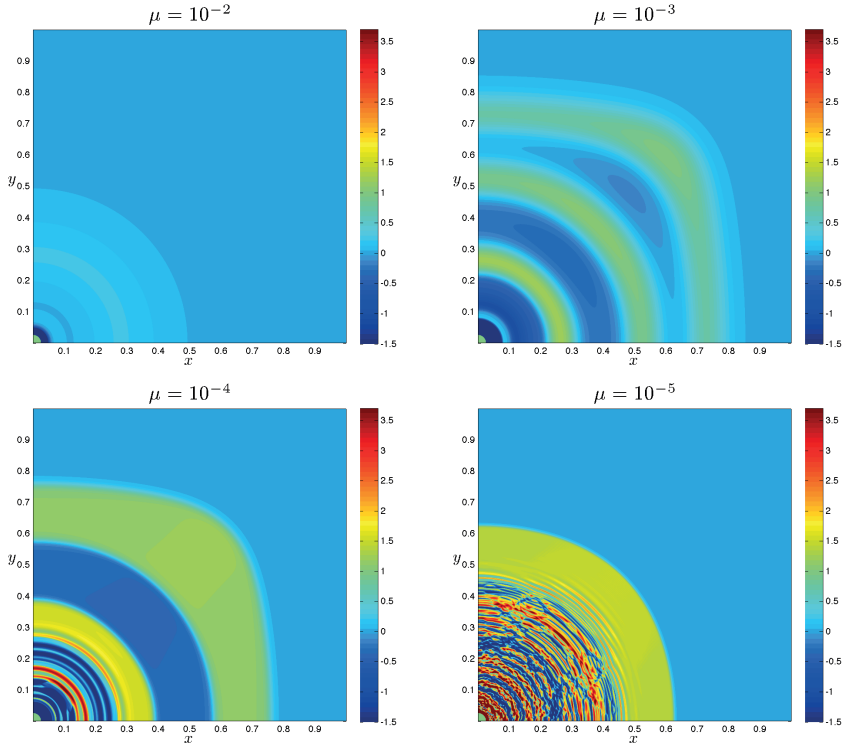

FIGURE 2. Numerical solutions of (2.1) using (4.4) – (4.29) on a 2000×2000 grid at $t = 0.65$.

system

$$\begin{cases} s_t^\mu + v_x^\mu = 0, \\ -\mu v_{xx}^\mu + v^\mu = f(s^\mu), \end{cases} \qquad \text{for } t > 0 \text{ and } x \in \mathbb{R}. \tag{5.1}$$

We look for a traveling wave solution to this system on the form

$$s^\mu(x,t) = s\left(\frac{x - \sigma t}{\sqrt{\mu}}\right), \quad v^\mu(x,t) = v\left(\frac{x - \sigma t}{\sqrt{\mu}}\right),$$

for some functions s and v. Inserting this into (5.1), we find

$$-\sigma s' + v' = 0, \quad -v'' + v = f(s).$$

We want to have

$$\lim_{\xi\to-\infty} s(\xi) = s_l, \quad \lim_{\xi\to\infty} s(\xi) = s_r \quad \text{and} \quad \lim_{|\xi|\to\infty} v''(\xi) = 0.$$

Thus the first equation can be integrated to get

$$-\sigma s + v = C, \quad C = f(s_l) - \sigma s_l = f(s_r) - \sigma s_r,$$

which yields the Rankine-Hugoniot condition

$$\sigma = \frac{f(s_r) - f(s_l)}{s_r - s_l}.$$

This means that any traveling wave will travel with a speed such that the limit $\lim_{\mu\to 0} s^\mu(x,t)$ is a weak solution to the conservation law (first equation in (5.1)). We are now left with the second order equation

$$-\sigma s'' + \sigma(s - s_l) = f(s) - f(s_l),$$

or equivalently, the system of first order equations

$$\begin{aligned} s' &= w, \\ w' &= (s - s_l) - \frac{1}{\sigma}\left(f(s) - f(s_l)\right). \end{aligned}$$

This system is integrable, and the solutions are the contour lines of

$$H(s,w) = \frac{\sigma}{2}w^2 - \frac{\sigma}{2}(s - s_l)^2 + \int_{s_l}^{s} (f(z) - f(s_l))\, dz.$$

Thus all fixed points are either stable centers or saddle points, located along the s-axis. Since $H_{ww} > 0$, the saddle points will be fixed points where $H_{ss} \le 0$, i.e.,

$$\sigma \ge f'(s). \tag{5.2}$$

The fixed points where $\sigma < f'(s)$ will be stable centers, and cannot be left or right states of traveling waves. Since $f(s)$ is assumed to be "*s-shaped*", for any s_l in $[0,1]$, except for the two values where f'' has extrema, there will be two other points s_1 and s_2 such that the Rankine-Hugoniot condition holds. Either one of the largest and the smallest of the three points s_l, s_1 and s_2 will be saddle points, and the middle point will be a center.

Also, independently of the shape of f, the condition (5.2) is necessary for a traveling wave. This means that the limits of such a traveling wave can only satisfy the Lax entropy condition, $f'(s_l) \ge \sigma \ge f'(s_r)$, if either $f'(s_l) = \sigma = f'(s_r)$ which means that f is linear between s_l and s_r or if $f'(s_l) = \sigma > f'(s_r)$. In the latter case however, the reverse shock wave, with s_l and s_r interchanged, is not Lax-admissible.

If the two saddle points are on the same contour line of H, there is a traveling wave connecting s_l with s_r, as well as its mirror image in the (s,w) plane, connecting s_r with

s_l. At least one of the two traveling waves cannot converge to an entropic shock as $\mu \to 0$ by the argument above. If there is a connecting orbit, then $H(s_l, 0) = H(s_r, 0)$, or

$$\frac{1}{2}\left(f(s_r) - f(s_l)\right)(s_r - s_l) = \int_{s_l}^{s_r} f(z) - f(s_l)\, dz. \tag{5.3}$$

Now let us assume that $1/2 - f(1/2 - \kappa) = f(1/2 + \kappa) - 1/2$, which is the case for the model flux function

$$f(s) = \frac{s^2}{s^2 + (1-s)^2}.$$

Then (5.3) implies that there is a traveling wave if and only if $|s_l - 1/2| = |s_r - 1/2|$. In particular, there is a traveling wave from $s = 0$ to $s = 1$ as well as one from $s = 1$ to $s = 0$.

There is substantial evidence that the numerical schemes also converge to this non-entropic traveling wave for small μ. In Figure 3 we show a computation using the simple finite difference scheme,

$$\begin{cases} s_j^{n+1} = s_j^n - \frac{\Delta t}{2\Delta x}\left(v_j^n - v_{j-1}^n\right), \\ -\frac{\mu}{\Delta x^2}\left(v_{j+1}^n - 2v_j^n + v_{j-1}^n\right) + v_j^n = f\left(s_j^n\right), \end{cases} \qquad \text{for } j \in 1, \ldots, N,\ n \geq 0, \tag{5.4}$$

where $\Delta x = 1/N$, and $\Delta t = 0.4\Delta x$. We used initial values

$$s_j^0 = \begin{cases} 1, & j\Delta x < 0.02, \\ 0, & \text{otherwise}, \end{cases}$$

and boundary values $v_0^n = 1$, $v_{N+1}^n = 0$ and $s_0^n = 1$, $s_{N+1}^n = 0$. The figure clearly shows that even for very small $\mu = 10^{-6}$, the solution is a traveling discontinuity that connects 1 and 0. On the other hand, the standard entropy solution for the limit conservation law ($\mu = 0$) is given by a wave connecting 1 to some intermediate state and a shock front between this intermediate state and 0.

Furthermore, we have also studied the possible convergence as $\mu \to 0$. In order to do this, we chose initial data which were not endpoints for the traveling wave solution. In Figure 4 we show the computed solutions at $t = 0.65$ using 10^4 mesh points in the interval $[0, 1]$ for three different values of μ. In this case the initial values were

$$s_0(x) = \begin{cases} 0.8, & x \leq 0.02, \\ 0.8\exp(-150(x - 0.02)^2), & \text{otherwise}. \end{cases} \tag{5.5}$$

From this figure, it seems that the limit (if any such limit exists) as $\mu \to 0$ of s^μ is not the entropy solution to the conservation law. This entropy solution is also indicated in Figure 4, and differs from s^μ. As $\mu \to 0$, the computed solution seems to converge to two traveling discontinuities, one from $s = 0.8$ to 1 followed by one from 1 to 0. Only the first of these is a classical shock wave.

We have also included a test where the initial data is periodic, viz.,

$$s(0, x) = \frac{1}{2}\left(1 + \cos(2\pi x)\right). \tag{5.6}$$

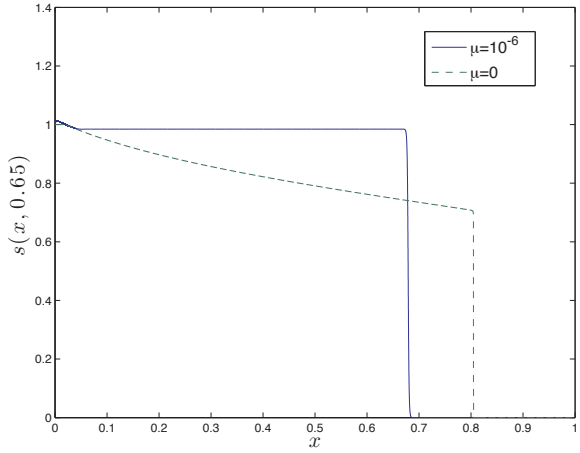

FIGURE 3. The numerical solution with $\mu = 10^{-6}$ and $\mu = 0$, and $N = 25\,000$.

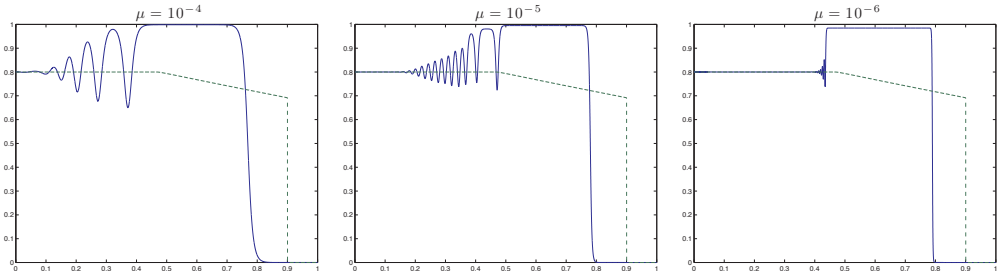

FIGURE 4. The computed solution to (5.5) at $t = 0.65$ for $\mu = 10^{-4}$ (left), $\mu = 10^{-5}$ (middle) and $\mu = 10^{-6}$ (right). In these computations, $N = 25\,000$.

In order to check the possible convergence as $\mu \to 0$, we computed approximations with $N = 25\,000$, and $t \in [0, 1]$. In Figure 5 we show the result in the (x, t) plane for $\mu = 10^{-6}$ and $\mu = 0$. The two solutions are almost identical until shocks develop at $t \approx 0.05$. At this point the approximation with $\mu = 10^{-6}$ develops two shocks, the slower (and weaker) is an entropy satisfying shock wave, while the faster (and stronger) violates the entropy condition. From the figure it is visible how the characteristics "pass through" the shock. Of course, if $\mu = 0$ the scheme reduces to the upwind scheme, and the approximation to the right is close to the entropy solution. The small entropic shock wave cannot be a traveling wave solution, whereas the large non-entropic shock wave is, since it is symmetric about $s = 1/2$. This follows from the previous analysis, and can be seen by the trailing oscillations in the small shock, these are absent in the large shock, see Figure 6.

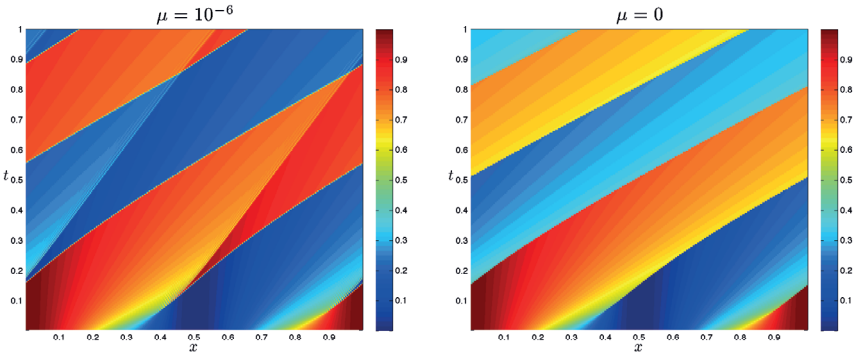

FIGURE 5. Approximations to the solution to (5.6) in the (x, t) plane, left: $\mu = 10^{-6}$, right: $\mu = 0$, $N = 25\,000$.

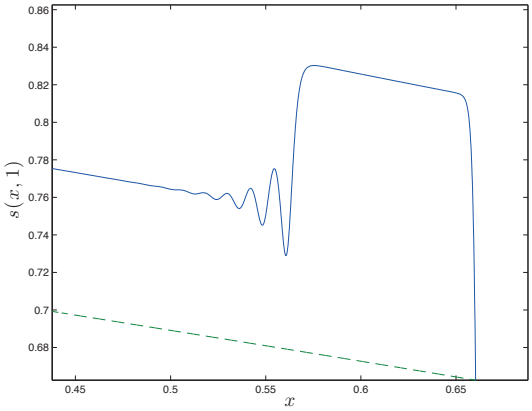

FIGURE 6. Trailing oscillations behind the entropic shock wave.

REMARK 5.1. The above simulations clearly indicate that the $\mu \to 0$ limit for the Brinkman regularization results in a non-classical shock (see [**90**] for definition) of the limit conservation law $(s_t + f(s)_x = 0)$. Such non-classical shocks in the context of two-phase flows in one-dimensional porous media also arise in the models with dynamic capillary pressure, see [**63, 67, 27, 81, 80**]. The existence of non-classical shocks for this model was proved in [**121, 120**]. It is interesting to observe that non-classical shock waves for two-phase flows can arise with two very different regularization mechanisms, one involving dynamic capillary pressure and one with a Brinkman regularization of the Darcy's law.

REMARK 5.2. We may also notice that system (5.1) is invariant with respect to the change of variable $x \to -x$, $t \to -t$, that is, the solution is reversible in time for any $\mu > 0$. This stands in contrast to the time irreversibility of entropy solutions to conservation laws such as $s_t + f(s)_x = 0$, to which (5.1) formally reduces as $\mu \to 0$.

5.0.3. *Convergence of the scheme in 1D.* In order to substantiate the above one-dimensional numerical calculations, we devote a short section to prove that the scheme (5.4) produces a convergent subsequence. We note that the scheme (5.4) is different from the two-dimensional finite difference scheme (4.4) – (4.9) for the two-dimensional case as no pressure equations are solved in the one-dimensional case.

For ease of notation, we write s and v rather than s^μ and v^μ. A solution to (5.1) is defined as a pair of functions (s, v) such that

$$s \in L^\infty(0,T;H^1(\mathbb{R})) \cap L^\infty((0,T)\times\mathbb{R}), \qquad v \in L^\infty(0,T;H^2(\mathbb{R})), \tag{5.7}$$

and such that for all test functions $\varphi \in C_0^\infty(\mathbb{R}\times[0,\infty))$,

$$\int_0^\infty \int_{\mathbb{R}} s\varphi_t + v\varphi_x \, dxdt + \int_{\mathbb{R}} s_0(x)\varphi(x,0)\, dx = 0, \tag{5.8}$$

$$\int_0^T \int_{\mathbb{R}} \mu v_x\varphi_x + v\varphi - f(s)\varphi \, dxdt = 0. \tag{5.9}$$

Using the obvious notation, (5.4) reads

$$\begin{cases} D_t^+ s_j^n + D_- v_j^n = 0, \\ -\mu D_+ D_- v_j^n + v_j^n = f(s_j^n). \end{cases} \qquad s_j^0 = s_0(j\Delta x), \text{ for } j \in \mathbb{Z}. \tag{5.10}$$

Let $v_{\Delta x}$ and $s_{\Delta x}$ be the piecewise constant functions defined by

$$\left.\begin{aligned} s_{\Delta x}(x,t) &= s_j^n, \\ v_{\Delta x}(x,t) &= v_j^n, \end{aligned}\right\} \qquad \text{for } (x,t) \in [x_{j-1/2}, x_{j+1/2}) \times [t_n, t_{n+1}).$$

As in the two dimensional case, we introduce the discrete norms

$$|v_{\Delta x}|_{h^1}^2 = \Delta x \sum_j \left|D_- v_j^n\right|^2,$$

$$\|v_{\Delta x}\|_{h^1}^2 = \|v_{\Delta x}\|_{L^2(\mathbb{R})} + |v_{\Delta x}|_{h^1}^2,$$

$$|v_{\Delta x}|_{h^2}^2 = \Delta x \sum_j \left|D_+ D_- v_j^n\right|^2,$$

$$\|v_{\Delta x}\|_{h^2}^2 = \|v_{\Delta x}\|_{h^1}^2 + |v_{\Delta x}|_{h^2}^2.$$

In order to show the strong convergence of a subsequence we square the equation for v_j^n and sum over j to find

$$\mu^2 \sum_j \left|D_- D_+ v_j^n\right|^2 + 2\mu \sum_j \left|D_- v_j^n\right|^2 + \sum_j \left|v_j^n\right|^2 = \sum_j \left|f_j^n\right|^2.$$

This means that

$$|v_{\Delta x}(\cdot,t)|_{h^2}^2 + |v_{\Delta x}(\cdot,t)|_{h^1}^2 + \|v_{\Delta x}(\cdot,t)\|_{L^2(\mathbb{R})}^2 \leq C\,\|f\|_{\mathrm{Lip}}^2\,\|s_{\Delta x}(\cdot,t)\|_{L^2(\mathbb{R})}^2\,, \tag{5.11}$$

for some constant C which does not depend on Δx. Next, we note that

$$\begin{aligned}\|s_{\Delta x}(\cdot,t_{n+1})\|_{L^2(\mathbb{R})} &\leq \|s_{\Delta x}(\cdot,t_n)\|_{L^2(\mathbb{R})} + C\Delta t\,|v_{\Delta x}(\cdot,t_n)|_{h^1}\\ &\leq \|s_{\Delta x}(\cdot,t_n)\|_{L^2(\mathbb{R})}\left(1 + C\Delta t\,\|f\|_{\mathrm{Lip}}\right).\end{aligned}$$

Thus

$$\|s_{\Delta x}(\cdot,t)\|_{L^2(\mathbb{R})} \leq \|s_0\|_{L^2(\mathbb{R})}\,e^{Ct},$$

for some constant C which does not depend on Δx (but scales like $1/\mu$). Combining this with (5.11) we find that

$$\|v_{\Delta x}(\cdot,t)\|_{h^2} \leq C_T$$

for all $t \leq T$. This means that we get a supremum bound on $s_{\Delta x}$, since

$$\sup_j \left|D_- v_j^n\right| \leq \|v_{\Delta x}(\cdot,t_n)\|_{h^2}\,.$$

Therefore

$$\|s_{\Delta x}(\cdot,t)\|_{L^\infty(\mathbb{R})} \leq \|s_0\|_{L^\infty(\mathbb{R})} + tC_T. \tag{5.12}$$

In particular, this implies that in the one dimensional case we can relax (**H.2**). Indeed, we only have to demand that f is *locally* Lipschitz continuous.

Now set $r_j^n = D_t^+ s_j^n$ and $z_j^n = D_t^+ v_j^n$. Then

$$\begin{cases} D_t^+ r_j^n + D_- z_j^n = 0,\\ -\mu D_+ D_- z_j^n + z_j^n = f'\left(s_j^{n+1/2}\right) r_j^n, \end{cases} \qquad n \geq 0,$$

where $s_j^{n+1/2}$ is some value between s_j^n and s_j^{n+1}. The above holds for $n \geq 0$, and we have that

$$r_j^0 = -D_- v_j^0, \text{ or } -\mu D_+ D_- r_j^0 + r_j^0 = -f'\left(\bar{s}_j^0\right) D_- s_j^0,$$

where $\bar{s}_j^0$ is a value between s_{j-1}^0 and s_j^0. Now we can repeat the above arguments to show that

$$\begin{aligned}\left\|D_t^+ v_{\Delta x}(\cdot,t)\right\|_{h^2} &\leq C\,\|f\|_{\mathrm{Lip}}\left\|D_t^+ s_{\Delta x}(\cdot,t)\right\|_{L^2(\mathbb{R})}\,,\\ \left\|D_t^+ s_{\Delta x}(\cdot,t)\right\|_{L^2(\mathbb{R})} &\leq \|f\|_{\mathrm{Lip}}\,\|s_0\|_{L^2(\mathbb{R})}\,e^{Ct}.\end{aligned}$$

Thus, if $s_0 \in H^1(\mathbb{R})$, then $D_t^+ s_{\Delta x} \in L^\infty(0,T;L^2(\mathbb{R}))$ and $D_t^+ v_{\Delta x} \in L^\infty(0,T;h^2)$, with bounds independent of Δx.

Now we need to show the compactness of the two sequences $\{s_{\Delta x}\}_{\Delta x>0}$ and $\{v_{\Delta x}\}_{\Delta x>0}$. Set $\sigma_j^n = D_- s_j^n$ and $w_j^n = D_- v_j^n$, then

$$\begin{cases} D_t^+ \sigma_j^n + D_- w_j^n = 0,\\ -\mu D_+ D_- w_j^n + w_j^n = f'\left(s_{j-1/2}^n\right)\sigma_j^n, \end{cases} \qquad n \geq 0,$$

where $s^n_{j-1/2}$ is an intermediate value. The initial values for the above scheme are $\sigma^0_j = D_- v^0_j$. From this we obtain

$$\|D_- v_{\Delta x}(\cdot,t)\|_{h^2} \le C\,\|f\|_{\mathrm{Lip}}\,|s_{\Delta x}|_{h^1}\,, \tag{5.13}$$

$$|s_{\Delta x}(\cdot,t)|_{h^1} \le \|f\|_{\mathrm{Lip}}\,\|\partial_x s_0\|_{L^2(\mathbb{R})}\,e^{Ct}. \tag{5.14}$$

Together with [**88**, Lemma 3.1, 3.2; Thm. 3.2, 4.1], these estimates imply

$$\begin{gathered} s_{\Delta x} \to s \text{ in } L^\infty(0,T;L^2(\mathbb{R})),\\ (v_{\Delta x}, D_- v_{\Delta x}, D^2_- v_{\Delta x}) \to (v, \partial_x v, \partial^2_{xx} v) \text{ in } L^\infty(0,T;L^2(\mathbb{R})),\\ D_- s_{\Delta x} \rightharpoonup \partial_x s, \text{ in } L^\ell(0,T;L^2(\mathbb{R})),\ 1 \le \ell < \infty, \text{ and}\\ D^3_- v_{\Delta x} \rightharpoonup \partial^3_{xxx} v, \text{ in } L^\ell(0,T;L^2(\mathbb{R})),\ 1 \le \ell < \infty, \end{gathered} \tag{5.15}$$

for a subsequence. Moreover, thanks to (5.12), we have that the limit $s \in L^\infty((0,T)\times\mathbb{R})$. Therefore since this scheme is conservative, it follows from standard arguments used e.g., in proving the Lax-Wendroff theorem, see [**73**], that the limits s and v satisfy (5.8) and (5.9) respectively. To sum up, we have proved

LEMMA 5.3. *Assume that $s_0 \in H^1(\mathbb{R})$ and that $s_{\Delta x}$ and $v_{\Delta x}$ are defined by* (5.10)*. Then there are functions $s \in L^\infty(0,T;H^1(\mathbb{R}))$ and $v \in L^\infty(0,T;H^3(\mathbb{R}))$ that are weak solutions to* (5.1)*, defined by* (5.7)*,* (5.8) *and* (5.9) *such that* (5.15) *holds.*

REMARK 5.4. As indicated by the above estimates, *for fixed $\mu > 0$*, the limit v is actually in $H^3(\mathbb{R})$ for each t, this means that $s_t \in H^2(\mathbb{R})$. Therefore we can use a bootstrap argument to show that both s and v are as regular as the initial data. Furthermore one easily shows the stability estimate

$$\|s(t,\cdot) - \bar{s}(t,\cdot)\|_{H^1(\mathbb{R})} \le C_\mu e^{t\|f\|_{\mathrm{Lip}}}\,\|s_0 - \bar{s}_0\|_{H^1(\mathbb{R})}\,,$$

where $\bar{s}$ is a solution with initial data $\bar{s}_0$. This shows that weak solutions are unique for each $\mu > 0$.

6. Conclusions

Under the assumption of vanishing capillary pressure, two-phase flows in a porous medium are modeled by a hyperbolic equation for the saturation, coupled with an elliptic equation for the pressure, resulting in the classical Darcy's law based equations (1.8). No existence results for the equations have been obtained till date in spite of the extensive research on these equations over the past several decades. One of the pressing issues in this context has been whether the Darcy's law is an adequate and appropriate model for flows in porous media. The Brinkman regularization of the Darcy's law [**17**] has been a popular alternative ([**84**] and references therein) for the Darcy's law in the geophysics community, at least in the context of a single phase flow. It is natural to examine whether the Brinkman regularization is an appropriate model, also in the context of two- (and multi-) phase flows in porous media.

In this paper, we consider the Brinkman regularization of the two-phase flow equations (1.10). A suitable notion of weak solutions for these equations is proposed. We prove that these weak solutions exist. Furthermore, a simple finite difference scheme to approximate this system (1.10) is proposed and is shown to converge to a weak solution. Numerical experiments indicate robust performance of this numerical scheme, for fixed Brinkmann regularization parameter μ.

Formally, we can recover the classical two-phase flow equations (1.8) by setting the regularization parameter $\mu \to 0$ in the Brinkman regularization (1.10). However, our stability estimates on the saturation and the velocity blow up as $\mu \to 0$, thus preventing us from proving that the limit solution of the Brinkman regularization is a weak solution of the classical Darcy problem. We investigate this question numerically using the convergent numerical scheme. Results on a benchmark quarter five-spot problem in two space dimensions show that the approximate solutions to the Brinkman regularization can become quite oscillatory as $\mu \to 0$. Furthermore, the regularized system can contain discontinuous fronts connecting full water saturation to zero water saturation. Such solutions are not included as classical entropy solutions of the Darcy problem (1.10). Hence, the numerical results indicate that the Brinkman regularization may not converge to (entropy solutions of) the Darcy limit as $\mu \to 0$.

This proposition is further investigated in the special case of one space dimension. In this case, the pressure equation is trivially solved and the saturation is modeled by a scalar conservation law. Entropy solutions (obeying Lax type entropy conditions) are widely recognized as the physically relevant solutions in this context. However, we establish using traveling wave analysis that the Brinkman limit will lead to a non-classical shock wave for the scalar conservation law. Such non-entropic solutions have been postulated for other physical models such as dynamic capillary pressure models [**63, 67**]. The presence of non-classical shocks for the Brinkman limit raises interesting questions, see also [**45**].

Summarizing, the Brinkman regularization does provide a model where existence of weak solutions can be shown rigorously and convergent numerical schemes can also be designed. Such existence and convergence results have not been possible for the Darcy problem despite several attempts. On the other hand, the Brinkman regularization may lead to limit solutions of the Darcy's equation that are not entropic and may contain non-classical shock waves. Furthermore, the question of rigorous passage to the Darcy limit for the Brinkman regularization is still wide open. Hence, this paper advocates caution in the use of Brinkman type models, at least for two and multi-phase flows in porous media.

An important assumption in the current paper has been that of zero capillary pressure. The inclusion of capillary pressure into our model and a study of the resultant effect of the Brinkman regularization will be performed in a future work.

7. Appendix

7.1. Inequalities. Let $\Omega = [0,1]^2$ be discretized by a grid with grid points $(x_i, y_j) = ((i-1/2)\Delta x, (j-1/2)\Delta y))$, $i,j = 0,\ldots,N+1$, $\Delta x = \Delta y = 1/N$ and f_{ij} a quantity given

at the grid points (x_i, y_j). We denote the piecewise constant interpolation f_Δ,

$$f_\Delta(x,y) = f_{ij}, \quad (x,y) \in [x_{i-1/2}, x_{i+1/2}) \times [y_{j-1/2}, y_{j+1/2}),$$

$i, j = 0, \ldots, N+1$ as in Section 4. Moreover, we assume that f_Δ satisfies the 'boundary conditions',

$$\begin{aligned} D^y_- f_{i1} = \Pi_{i,1/2}, \quad D^y_+ f_{iN} = \Pi_{i,N+1/2}, \quad i = 1, \ldots, N, \\ D^x_- f_{1j} = \Pi_{1/2,j}, \quad D^x_+ f_{Nj} = \Pi_{N+1/2,j}, \quad j = 1, \ldots, N. \end{aligned}$$

Then the following lemma holds:

LEMMA 7.1. *(Discrete trace inequality)*

$$\begin{aligned} (7.1) \quad \|f_\Delta\|^2_{L^2(\partial_\Delta)} \leq 2\|f_\Delta\|^2_{H^1_\Delta} + 6\|f_\Delta\|^2_{L^2(\Omega)} \\ + 2\Delta y \Delta x \left\{ \sum_{j=1}^N \left(\Pi_{N+1/2,j}\right)^2 + \sum_{i=1}^N \left(\Pi_{i,N+1/2}\right)^2 \right\}, \end{aligned}$$

where $\Pi_{N+1/2,j} = D^x_+ f_{Nj}$, $\Pi_{i,N+1/2} = D^y_+ f_{iN}$ *are the Neumann boundary conditions, where the norms are defined in* (4.10) *and* (4.11).

PROOF. First, we note that we can write

$$\begin{aligned} f_{N+1,j} &= \sum_{i=1}^N \left(\frac{i}{N} f_{i+1j} - \frac{i-1}{N} f_{ij} \right) \\ &= \sum_{i=1}^N \left(\frac{i}{N} (f_{i+1j} - f_{ij}) + \frac{1}{N} f_{ij} \right) \\ &= \Delta x \sum_{i=1}^N \left(\frac{i}{N} D^x_+ f_{ij} + f_{ij} \right) \\ &= \Delta x \sum_{i=1}^N \left(\frac{i-1}{N} D^x_- f_{ij} + f_{ij} \right) + \Delta x\, \Pi_{N+1/2,j}, \end{aligned}$$

and similarly,

$$\begin{aligned} f_{i,N+1} &= \Delta y \sum_{j=1}^N \left(\frac{j-1}{N} D^y_- f_{ij} + f_{ij} \right) + \Delta y\, \Pi_{i,N+1/2}, \\ f_{0j} &= -\Delta x \sum_{i=1}^N \left(\frac{N+1-i}{N} D^x_- f_{ij} - f_{ij} \right), \\ f_{i0} &= -\Delta y \sum_{j=1}^N \left(\frac{N+1-j}{N} D^y_- f_{ij} - f_{ij} \right). \end{aligned}$$

Thus,

$$
\begin{aligned}
\Delta y \sum_{j=1}^{N} (f_{N+1,j})^2 &= \Delta y \sum_{j=1}^{N} \left(\Delta x \sum_{i=1}^{N} \left(\frac{i-1}{N} D_-^x f_{ij} + f_{ij} \right) + \Delta x \, \Pi_{N+1/2,j} \right)^2 \\
&= \Delta y \Delta x^2 \sum_{j=1}^{N} \left(\sum_{i=1}^{N} \left(\frac{i}{N} D_+^x f_{ij} + f_{ij} \right) \right)^2 \\
&\leq \Delta y \Delta x \sum_{i,j=1}^{N} \left(\frac{i}{N} D_+^x f_{ij} + f_{ij} \right)^2 \\
&\leq 2 \Delta y \Delta x \sum_{i,j=1}^{N} \left(\left(\frac{i}{N} D_+^x f_{ij} \right)^2 + (f_{ij})^2 \right) \\
&= 2 \Delta y \Delta x \sum_{i,j=1}^{N} \left(\frac{i-1}{N} D_-^x f_{ij} \right)^2 + 2 \Delta y \Delta x \sum_{i,j=1}^{N} (f_{ij})^2 \\
&\quad + 2 \Delta y \Delta x \sum_{j=1}^{N} \left(\Pi_{N+1/2,j} \right)^2 ,
\end{aligned} \tag{7.2}
$$

and in the same way,

$$
\begin{aligned}
\Delta x \sum_{i=1}^{N} (f_{i,N+1})^2 &\leq 2 \Delta y \Delta x \sum_{i,j=1}^{N} \left(\frac{j-1}{N} D_-^y f_{ij} \right)^2 + 2 \Delta y \Delta x \sum_{i,j=1}^{N} (f_{ij})^2 \\
&\quad + 2 \Delta y \Delta x \sum_{i=1}^{N} \left(\Pi_{i,N+1/2} \right)^2 , \\
\Delta y \sum_{j=1}^{N} (f_{0j})^2 &\leq 2 \Delta y \Delta x \sum_{i,j=1}^{N} \left(\frac{N+1-i}{N} D_-^x f_{ij} \right)^2 + 2 \Delta y \Delta x \sum_{i,j=1}^{N} (f_{ij})^2 \\
\Delta x \sum_{i=1}^{N} (f_{i0})^2 &\leq 2 \Delta y \Delta x \sum_{i,j=1}^{N} \left(\frac{N+1-j}{N} D_-^y f_{ij} \right)^2 + 2 \Delta y \Delta x \sum_{i,j=1}^{N} (f_{ij})^2 .
\end{aligned} \tag{7.3}
$$

We note that

$$
\left(\frac{k-1}{N} \right)^2 + \left(\frac{N+1-k}{N} \right)^2 \leq 1,
$$

hence summing up (7.2) – (7.3), we find (7.1). □

Moreover, we have the following discrete version of the Poincaré inequality:

LEMMA 7.2. *(Discrete Poincaré inequality) Denote* $\overline{f} := \Delta x \Delta y \sum_{i,j=1}^{N} f_{ij}$. *Then we have*

$$\|f_\Delta - \overline{f}\|^2_{L^2(\Omega)} \leq \frac{8}{9}\Delta x \Delta y \sum_{i,j=1}^{N} \left(|D^x_- f_{ij}|^2 + |D^y_- f_{ij}|^2 \right). \tag{7.4}$$

PROOF. (C.S:= Cauchy Schwarz inequality, Y. := Young's inequality)

$$\begin{aligned}
\|f_\Delta - \overline{f}\|^2_{L^2(\Omega)} &:= \Delta x \Delta y \sum_{i,j=1}^{N} \left(f_{ij} - \overline{f} \right)^2 \\
&= \Delta x \Delta y \sum_{i,j=1}^{N} \left(f_{ij} - \Delta x \Delta y \sum_{k,\ell=1}^{N} f_{k\ell} \right)^2 \\
&= \Delta x \Delta y \sum_{i,j=1}^{N} \left(\Delta x \Delta y \sum_{k,\ell=1}^{N} (f_{ij} - f_{k\ell}) \right)^2 \\
&= \Delta x \Delta y \sum_{i,j=1}^{N} \left(\Delta x \Delta y \sum_{k,\ell=1}^{N} (f_{ij} - f_{kj} + f_{kj} - f_{k\ell}) \right)^2 \\
&= \Delta x^3 \Delta y^3 \sum_{i,j=1}^{N} \left(\sum_{k,\ell=1}^{N} \left(\Delta x \sum_{m=i+1}^{k} D^x_- f_{mj} + \Delta y \sum_{n=j+1}^{\ell} D^y_- f_{kn} \right) \right)^2 \\
&\overset{\text{C.S.}}{\leq} \Delta x^3 \Delta y^3 \sum_{i,j=1}^{N} \left(\sum_{k,\ell=1}^{N} \left(\Delta x \sqrt{|k-i|} \left(\sum_{m=1}^{N} |D^x_- f_{mj}|^2 \right)^{\frac{1}{2}} \right.\right. \\
&\qquad \left.\left. + \Delta y \sqrt{|\ell - j|} \left(\sum_{n=1}^{N} |D^y_- f_{kn}|^2 \right)^{\frac{1}{2}} \right) \right)^2 \\
&= \Delta x \Delta y \sum_{i,j=1}^{N} \left(\Delta x^2 \sum_{k=1}^{N} \sqrt{|k-i|} \left(\sum_{m=1}^{N} |D^x_- f_{mj}|^2 \right)^{\frac{1}{2}} \right. \\
&\qquad \left. + \Delta y^2 \Delta x \sum_{\ell=1}^{N} \sqrt{|\ell - j|} \sum_{k=1}^{N} \left(\sum_{n=1}^{N} |D^y_- f_{kn}|^2 \right)^{\frac{1}{2}} \right)^2 \\
&\leq \Delta x \Delta y \sum_{i,j=1}^{N} \left(\frac{2}{3} \left(\Delta x \sum_{m=1}^{N} |D^x_- f_{mj}|^2 \right)^{\frac{1}{2}} \right. \\
&\qquad \left. + \frac{2}{3} \Delta x \sum_{k=1}^{N} \left(\Delta y \sum_{n=1}^{N} |D^y_- f_{kn}|^2 \right)^{\frac{1}{2}} \right)^2
\end{aligned}$$

$$\overset{\text{Y.}}{\leq} \frac{8}{9}\Delta x\Delta y \sum_{i,j=1}^{N}\left[\Delta x \sum_{m=1}^{N}|D_-^x f_{mj}|^2 + \left(\Delta x \sum_{k=1}^{N}\left(\Delta y \sum_{n=1}^{N}|D_-^y f_{kn}|^2\right)^{\frac{1}{2}}\right)^2\right]$$

$$\overset{\text{Y.}}{\leq} \frac{8}{9}\Delta x\Delta y \sum_{j,m=1}^{N}|D_-^x f_{mj}|^2 + \frac{8}{9}\Delta x\Delta y \sum_{k,n=1}^{N}|D_-^y f_{kn}|^2$$

$$= \frac{8}{9}\Delta x\Delta y \sum_{i,j=1}^{N}\left(|D_-^x f_{ij}|^2 + |D_-^y f_{ij}|^2\right).$$

□

PAPER 3

A Convergent Explicit Finite Difference Scheme for a Mechanical Model for Tumor Growth

Joint work with Konstantina Trivisa

ABSTRACT. Mechanical models for tumor growth have been used extensively in recent years for the analysis of medical observations and for the prediction of cancer evolution based on imaging analysis. This work deals with the numerical approximation of a mechanical model for tumor growth and the analysis of its dynamics. The system under investigation is given by a multi-phase flow model: The densities of the different cells are governed by a transport equation for the evolution of tumor cells, whereas the velocity field is given by a Brinkman regularization of the classical Darcy's law. An efficient finite difference scheme is proposed and shown to converge to a weak solution of the system. Our approach relies on convergence and compactness arguments in the spirit of Lions [**93**].

1. Introduction

1.1. Motivation. Mechanical models for tumor growth are used extensively in recent years for the prediction of cancer evolution based on imaging analysis. Such models are based on the assumption that the growth of the tumor is mainly limited by the competition for space. Mathematical modeling, analysis and numerical simulations together with experimental and clinical observations are essential components in the effort to enhance our understanding of the cancer development. The goal of this article is to make a further step in the investigation of such models by presenting a convergent explicit finite difference scheme for the numerical approximation of a Hele-Shaw-type model for tumor growth and by providing its detailed mathematical analysis. Even though the main focus in the present work is on the investigation of the evolution of the proliferating cells, it provides a mathematical framework that can potentially accommodate more complex systems that account for the presence of nutrient and drug application. This will be the subject of future investigation.

1.2. Governing equations. In the present context the tissue is considered as a multiphase fluid and the ability of the tumor to expand into a host tissue is then primarily driven by the cell division rate which depends on the local cell density and the mechanical pressure in the tumor.

1.2.1. *Transport equations for the evolution of the cell densities.* The dynamics of the cell population density $n(t, x)$ under pressure forces and cell multiplication is described by

a transport equation

$$\partial_t n - \operatorname{div}(n\boldsymbol{u}) = n\boldsymbol{G}(p), \quad x \in \Omega,\ t \geq 0 \tag{1.1}$$

where n represents the number density of tumor cells, $\boldsymbol{u}$ the velocity field and p the pressure of the *tumor*. Ω is a bounded domain in $\mathbb{R}^d$, $d = 2, 3$. The pressure law is given by

$$p(n) = an^\gamma, \tag{1.2}$$

where $\gamma \geq 2$. Following [**20, 108**], we assume that growth is directly related to the pressure through a function $\boldsymbol{G}(\cdot)$ which satisfies

$$\boldsymbol{G} \in C^1(\mathbb{R}), \quad \boldsymbol{G}'(\cdot) \leq -\beta < 0, \quad \boldsymbol{G}(P_M) = 0 \quad \text{for some } P_M > 0. \tag{1.3}$$

The pressure P_M is usually called *homeostatic pressure.* Here, and in what follows, for simplicity we let

$$\boldsymbol{G}(p) = \alpha - \beta p^\theta, \tag{1.4}$$

for some $\alpha, \beta, \theta > 0$.

1.2.2. *The tumor tissue as a porous medium.* The continuous motion of cells within the tumor region, typically due to proliferation, is represented by the velocity field $\boldsymbol{u} := \nabla W$ given by an alternative to Darcy's equation known as *Brinkman's equation*

$$p = W - \mu \Delta W \tag{1.5}$$

where μ is a positive constant describing the viscous like properties of tumor cells and p is the pressure given by (1.2).

Relation (1.5) consists of two terms. The first term is the usual Darcy's law, which in the present setting describes the tendency of cells to move down pressure gradients and results from the friction of the tumor cells with the extracellular matrix. The second term, on the other hand, is a dissipative force density (analogous to the Laplacian term that appears in the Navier-Stokes equation) and results from the internal cell friction due to cell volume changes. A second interpretation of relation (1.5) is the tumor tissue can be viewed as "fluid like." In other words, the tumor cells flow through the fixed extracellular matrix like a flow through a porous medium, obeying Brinkman's law.

The resulting model, governed by the transport equation (1.1) for the population density of cells, the elliptic equation (1.5) for the velocity field and a state equation for the pressure law (1.2), now reads

$$\begin{cases} \partial_t n - \operatorname{div}(n\nabla W) = \alpha n - \beta n^{\gamma\theta+1}, \quad x \in \Omega,\ t \geq 0 \\ -\mu\Delta W + W = an^\gamma. \end{cases} \tag{1.6}$$

We complete the system (1.6) with a family of initial data n_0 satisfying (for some constant C)

$$n_0 \geq 0, \quad p(n_0) \leq P_M, \quad \|n_0\|_{L^1(\mathbb{R}^d)} \leq C. \tag{1.7}$$

The objective of this work is to establish the global existence of weak solutions to the nonlinear model for tumor growth (1.6) by designing an efficient numerical scheme for its approximation and by showing that this scheme converges when the mesh is refined. The main ingredients of our approach and contribution to the existing theory include:

(1) The introduction of a suitable notion of solutions to the nonlinear system (1.6) consisting of the transport equation (1.1) and the Brinkman regularization (1.5).
(2) The construction of an approximating procedure which relies on an artificial vanishing viscosity approximation and the establishment of the suitable compactness in order to pass into the limit and to conclude convergence to the original system (cf. Section 3, Lemma 3.7).
(3) The design of an efficient numerical scheme for the numerical approximation of the nonlinear system (1.1)-(1.5).
(4) The proof of the convergence of the numerical scheme. In the center of the analysis lies the proof of the strong convergence of the cell densities. This is achieved by establishing the weak continuity of the *effective viscous pressure* in the spirit of Lions [**93**] (cf. Section 4, Lemma 4.8).
(5) The design of numerical experiments in order to establish that the finite difference scheme is effective in computing approximate solutions to the nonlinear system (1.6) (cf. Section 5).

For relevant results on the analysis and the numerical approximation of a two-phase flow model in porous media we refer the reader to [**31**]. Related results on the numerical approximation of compressible fluids employing the weak compactness tools developed by of Lions [**93**] in the discrete setting have been established by Karper *et al.* [**78, 74, 75, 76**] and Gallouët *et al.* [**49**].

Relevant work on the mathematical analysis of mechanical models of Hele-Shaw-type have been presented by Perthame *et al.* [**104, 105, 106, 107**]. The analysis in [**106**] establishes the existence of traveling wave solutions of the Hele-Shaw model of tumor growth with nutrient and presents numerical observations in two space dimensions. The present article is according to our knowledge the first article presenting rigorous analytical results on the global existence of general weak solutions to Hele-Shaw-type systems.

A different approach yielding results on the global existence of weak solutions to a nonlinear model for tumor growth in a general moving domain $\Omega_t \subset \mathbb{R}^3$ without any symmetry assumption and for finite large initial data is presented in [**41, 39, 40**]. But in contrast to the present nonlinear system, the transport equation for the evolution of cancerous cells in [**41, 40**] has a source term which is linear with respect to cell density.

Relevant results on nonlinear models for tumor growth governed by the Darcy's law for the evolution of the velocity field are presented by Zhao [**128**] based on the farmework introduced by Friedman *et al.* [**51, 26**].

1.3. Outline. The paper is organized as follows: Section 1 presents the motivation, modeling and introduces the necessary preliminary material. Section 2 provides a weak formulation of the problem and states the main result. Section 3 is devoted to the global existence of solutions via a vanishing viscosity approximation. In Section 4 we present an

efficient finite difference scheme for the approximation of the weak solution to system (1.6) on rectangular domains and Section 5 is devoted to numerical experiments. A discretized Aubin-Lions lemma and some technical lemmas are presented in Appendices A and B respectively.

2. Weak formulation and main results

NOTATION 2.1. For $\varphi : (0,T)\times\Omega \to \mathbb{R}$, $\boldsymbol{\varphi} : (0,T)\times\Omega \to \mathbb{R}^d$, we will denote by $\nabla\varphi := \nabla_x\varphi = (\partial_{x_1}\varphi, \dots, \partial_{x_d}\varphi)$ and $\operatorname{div}\boldsymbol{\varphi} := \operatorname{div}_x\boldsymbol{\varphi} = \sum_{i=1}^d \partial_{x_i}\boldsymbol{\varphi}^{(i)}$ the gradient and divergence in the spatial direction in Ω.

2.1. Weak solutions.

DEFINITION 2.2. Let Ω a bounded domain in $\mathbb{R}^d$, $d = 2, 3$, which is either rectangular or has a smooth boundary $\partial\Omega$ and $T > 0$ a finite time horizon. We say that (n, W, p) is a weak solution of problem (1.1)-(1.5) supplemented with initial data (n_0, W_0, p_0) satisfying (1.7) provided that the following hold:

• $(n, W, p) \geq 0$ represents a weak solution of (1.1)-(1.5) on $(0,T)\times\Omega$, i.e., for any test function $\varphi \in C_c^\infty([0,T]\times\mathbb{R}^d), T > 0$, the following integral relations hold

$$\int_{\mathbb{R}^d} n\varphi(\tau,\cdot)\,dx - \int_{\mathbb{R}^d} n_0\varphi(0,\cdot)dx = \int_0^\tau\int_{\mathbb{R}^d} (n\partial_t\varphi - n\nabla W\cdot\nabla\varphi + n\boldsymbol{G}(p)\varphi(t,\cdot))\,dxdt. \tag{2.1}$$

In particular,

$$n \in L^p((0,T)\times\Omega), \text{ for all } p \geq 1.$$

We remark that in the weak formulation, it is convenient that the equations (1.1) hold in the whole space $\mathbb{R}^d$ provided that the densities n are extended to be zero outside the tumor domain.

• Brinkman's equation (1.5) holds in the sense of distributions, i.e., for any test function $\varphi \in C_c^\infty(\mathbb{R}^d)$ satisfying

$$\varphi|_{\partial\Omega} = 0 \text{ for any } t \in [0,T],$$

the following integral relation holds for a.e. $t \in [0,T]$,

$$\int_\Omega an^\gamma\varphi\,dx = \int_\Omega \Big(\mu\nabla W\cdot\nabla\varphi + W\varphi\Big)dx. \tag{2.2}$$

and $p = n^\gamma$ almost everywhere. All quantities in (2.2) are required to be integrable, and in particular, $W \in L^\infty([0,T]; H^2(\Omega))$.

The main result of the article now follows.

THEOREM 2.3. *Let $\Omega \subset \mathbb{R}^d$ be a bounded domain with smooth boundary $\partial\Omega$, $0 < T < \infty$. Assume that the initial data $n_0 \in L^\infty(\Omega)$ with $0 \leq n_0 \leq n_\infty := P_M^{1/\gamma}$ and that $\boldsymbol{G}(\cdot)$ is of the form* (1.4). *Then the problem* (1.1)-(1.5)*, admits a weak solution in the sense specified in Definition* 2.2.

The following two remarks are now in order.

REMARK 2.4. In Section 3, such a solution is obtained as the limit of the vanishing viscosity approximations $(n_\varepsilon, W_\varepsilon, p_\varepsilon)$ of (3.1) to (1.6) as $\varepsilon \to 0$.

REMARK 2.5. In Section 4, such a solution is obtained in the case of a rectangular domain, as the limit of the sequence of approximations (n_h, W_h, p_h) computed by the numerical scheme (4.1) – (4.3) as $h \to 0$.

3. Global existence via vanishing viscosity

In this section we prove Theorem 2.3 by constructing an approximating scheme which relies on the addition of an artificial vanishing viscosity approximation

$$\begin{cases} \partial_t n_\varepsilon - \operatorname{div}(n_\varepsilon \nabla W_\varepsilon) = \alpha n_\varepsilon - \beta n_\varepsilon^{\gamma+1} + \varepsilon \Delta n_\varepsilon, \quad x \in \Omega,\ t \geq 0 \\ \mu \Delta W_\varepsilon - W_\varepsilon = a n_\varepsilon^\gamma, \\ n_\varepsilon(0,\cdot) = n_0^\varepsilon, \end{cases} \tag{3.1}$$

where n_0^ε is a smoothnened version of n_0, that is $n_0^\varepsilon = n_0 * \varphi_\varepsilon$ for a smooth function φ_ε with compact support, and a bounded domain $\Omega \in \mathbb{R}^d$ with smooth boundary or alternatively the d-dimensional torus $\mathbb{T}^d$, and we establish its convergence to the nonlinear system (1.6) at the continuous level. For simplicity, we assume $a = 1$ and homogeneous Neumann boundary conditions for n_ε and W_ε (if the domain is a torus $\mathbb{T}^d$ we can also use periodic boundary conditions).

THEOREM 3.1. *For every $\varepsilon > 0$, the parabolic-elliptic system* (3.1) *admits a unique smooth solution $(n_\varepsilon, W_\varepsilon, p_\varepsilon)$.*

PROOF. The proof of this result relies on classical arguments (cf. Ladyzhenskaya [**87**]), namely by employing the Contraction Mapping Principle and the regularity of the initial data one can show the existence of a unique solution $(n_\varepsilon, W_\varepsilon, p_\varepsilon)$ defined for a small time $T > 0$. Then one derives apriori estimates establishing that the solution does not blow up and in fact is defined for every time. Finally, a bootstrap argument yields the smoothness of the solution. □

The remaining part of this section aims to establish the necessary compactness of the approximate sequence of solutions $(n_\varepsilon, W_\varepsilon, p_\varepsilon)$.

3.1. A priori estimates. We start by proving that n_ε are uniformly bounded independent of $\varepsilon > 0$ and nonnegative:

LEMMA 3.2. *If $0 \leq n_\varepsilon(0,\cdot) \leq n_\infty := P_M^{1/\gamma} < \infty$ uniformly in $\varepsilon > 0$, then for any $t > 0$, the functions $n_\varepsilon(t,\cdot)$ are uniformly (in $\varepsilon > 0$) bounded and nonnegative, specifically,*

$$0 \leq \min_{(t,x)} n_\varepsilon(t,x) \leq \max_{(t,x)} n_\varepsilon(t,x) \leq n_\infty.$$

PROOF. First we notice that if W_ε has a maximum at a point x_0, then $\Delta W_\varepsilon(\cdot, x_0) \leq 0$ and therefore $W_\varepsilon = p_\varepsilon + \mu \Delta W_\varepsilon \leq p_\varepsilon$. Similarly, if it has a minimum at a point x_0, it will satisfy $\Delta W_\varepsilon(\cdot, x_0) \geq 0$ and therefore $W_\varepsilon \geq p_\varepsilon$. If W_ε attains a strict maximum on the boundary, i.e., there is a point $x_0 \in \partial\Omega$ such that $W_\varepsilon(x_0) > W_\varepsilon(x)$ for any other $x \in \Omega$, we

apply Hopf's Lemma, e.g. [**46**, p. 347], to the function $v := W_\varepsilon - \max_{(t,x)} p_\varepsilon(t,x)$ which satisfies

$$-\mu\Delta v + v = p_\varepsilon - \max_{(t,x)} p_\varepsilon(t,x) \le 0,$$

which has a strict maximum at the point x_0. If $v(x_0) \le 0$, then $W_\varepsilon \le W_\varepsilon(x_0) \le \max_{(t,x)} p_\varepsilon(t,x)$ and otherwise Hopf lemma gives $\nabla W_\varepsilon(x_0)\cdot\nu = \nabla v(x_0)\cdot\nu > 0$ where we have denoted the boundary normal ν, this contradicts the homogeneous boundary conditions. In a similar way we show that $W_\varepsilon \ge \min_{(t,x)} p_\varepsilon(t,x)$ (applying Hopf's lemma to $-W_\varepsilon$ and hence

$$\min_{(t,x)} p_\varepsilon(t,x) \le W_\varepsilon \le \max_{(t,x)} p_\varepsilon(t,x). \tag{3.2}$$

We rewrite the evolution equation for n_ε using the equation for the potential W_ε,

$$\partial_t n_\varepsilon - \nabla W_\varepsilon\cdot\nabla n_\varepsilon = n_\varepsilon \boldsymbol{G}(p_\varepsilon) + \frac{1}{\mu} n_\varepsilon(p_\varepsilon - W_\varepsilon) + \varepsilon\Delta n_\varepsilon. \tag{3.3}$$

Now assume (t_0,x_0) is a point, where $n_\varepsilon(t_0,x_0) \ge n_\infty$ reaches its maximum (and therefore also $p_\varepsilon(t_0,x_0) \ge P_M$ reaches a maximum). Then $\nabla n_\varepsilon(t_0,x_0) = 0$ and $\Delta n_\varepsilon(t_0,x_0) \le 0$. Hence

$$\partial_t n_\varepsilon(t_0,x_0) \le n_\varepsilon \boldsymbol{G}(p_\varepsilon) + \frac{1}{\mu} n_\varepsilon(p_\varepsilon - W_\varepsilon).$$

By (3.2), the second term on the right hand side is nonpositive and since $\boldsymbol{G}(p_\varepsilon(t_0,x_0)) \le 0$ for $p_\varepsilon \ge P_M$, we get

$$\partial_t n_\varepsilon(t_0,x_0) \le 0.$$

Hence n_ε will decrease and if initially $n_0 \le n_\infty$, this implies that $n_\varepsilon(t,\cdot) \le n_\infty$ for any later time $t \ge 0$. To show the nonnegativity of n_ε, we integrate the evolution equation for n_ε,

$$\frac{d}{dt}\int_\Omega n_\varepsilon dx = \int_\Omega n_\varepsilon \boldsymbol{G}(p_\varepsilon) dx.$$

On the other hand, multiplying the same equation by a regularized version of the sign function, integrating and then passing to the limit in the approximation, we have

$$\frac{d}{dt}\int_\Omega |n_\varepsilon| dx \le \int_\Omega |n_\varepsilon| \boldsymbol{G}(p_\varepsilon) dx,$$

Subtracting the two equations from one another, and using that $|n_\varepsilon| - n_\varepsilon \ge 0$,

$$\begin{aligned}\frac{d}{dt}\int_\Omega \big||n_\varepsilon| - n_\varepsilon\big| dx &\le \int_\Omega \big||n_\varepsilon| - n_\varepsilon\big| \boldsymbol{G}(p_\varepsilon) dx,\\ &\le \max_{s\in[0,P_M]} |\boldsymbol{G}(s)| \int_\Omega \big||n_\varepsilon| - n_\varepsilon\big| dx.\end{aligned}$$

Now using Grönwall's inequality and that $|n_0| - n_0 \equiv 0$ by assumption, we obtain

$$\int_\Omega \big||n_\varepsilon| - n_\varepsilon\big|(t) dx = 0$$

and thus that $n_\varepsilon(t,x) \ge 0$ almost everywhere. □

Next we prove a simple lemma on the regularity of W_ε.

LEMMA 3.3. *We have that*

$$W_\varepsilon \subset L^\infty([0,T];W^{2,q}(\Omega)),$$

for any $q \in [1,\infty)$ *uniformly in* $\varepsilon > 0$ *and*

$$W_\varepsilon, \Delta W_\varepsilon \subset L^\infty((0,T)\times\Omega)),$$

uniformly in $\varepsilon > 0$ *as well.*

PROOF. We square the equation for W_ε and integrate it over the spatial domain and then use integration by parts,

$$\begin{aligned}\int_\Omega |p_\varepsilon|^2 dx &= \int_\Omega |W_\varepsilon|^2 - 2\mu W_\varepsilon \Delta W_\varepsilon + \mu^2 |\Delta W_\varepsilon|^2 dx \\ &= \int_\Omega |W_\varepsilon|^2 + 2\mu |\nabla W_\varepsilon|^2 + \mu^2 |\nabla^2 W_\varepsilon|^2 dx.\end{aligned}$$

By the previous Lemma 3.2, we have that p_ε is uniformly bounded in $\varepsilon > 0$ and therefore that the left hand side of the above equation is bounded and that $W_\varepsilon \in L^\infty([0,T];H^2(\Omega))$. Using a Calderon-Zygmund inequality (e.g. [**53**, Thm. 9.11.]), we obtain $W_\varepsilon \in L^\infty([0,T];W^{2,q}(\Omega))$ for all $q \in [1,\infty)$. By the Sobolev embedding theorem, this implies that in particular $\nabla W_\varepsilon \in L^\infty((0,T)\times\Omega)$. The second claim follows from (3.2) and the uniform bound on the pressure proved in Lemma 3.2. □

3.2. Entropy inequalities for n_ε. To prove strong convergence of the approximating sequence $\{(n_\varepsilon, W_\varepsilon, p_\varepsilon)\}_{\varepsilon>0}$, it will be useful to derive entropy inequalities for n_ε. To this end, the following lemma will be useful:

LEMMA 3.4. *Let* $f : \mathbb{R} \to \mathbb{R}$ *be a smooth convex, nonnegative function and denote* $f_\varepsilon := f(n_\varepsilon)$. *Then* f_ε *satisfies the following identity*

$$\begin{aligned}\partial_t f_\varepsilon - \operatorname{div}(f_\varepsilon \nabla W_\varepsilon) - \varepsilon \Delta f(n_\varepsilon) \\ = (f'(n_\varepsilon)n_\varepsilon - f_\varepsilon)\Delta W_\varepsilon + f'(n_\varepsilon)n_\varepsilon \boldsymbol{G}(p_\varepsilon) - \varepsilon f''(n_\varepsilon)|\nabla n_\varepsilon|^2\end{aligned} \tag{3.4}$$

where

$$\varepsilon \int_0^T \int_\Omega f''(n_\varepsilon)|\nabla n_\varepsilon|^2 \, dxdt \le C, \tag{3.5}$$

with $C > 0$ *a constant independent of* $\varepsilon > 0$. *In particular, this implies that* $\partial_t f_\varepsilon = g_\varepsilon + k_\varepsilon$ *with* $g_\varepsilon \in L^1([0,T]\times\Omega)$ *and* $k_\varepsilon \in L^1([0,T];W^{-1,2}(\Omega))$.

PROOF. The identity (3.4) follows after multiplying the evolution equation for n_ε, (3.3), by $f'(n_\varepsilon)$ and using chain rule. Integrating the inequality in space and time, we obtain

$$\int_\Omega f_\varepsilon(T)\, dx + \varepsilon \int_0^T \int_\Omega f''(n_\varepsilon)|\nabla n_\varepsilon|^2 \, dxdt$$

$$= \int_\Omega f_\varepsilon(0)\,dx + \int_0^T\int_\Omega (f'(n_\varepsilon)n_\varepsilon - f_\varepsilon)\Delta W_\varepsilon + f'(n_\varepsilon)n_\varepsilon \boldsymbol{G}(p_\varepsilon)\,dxdt$$

The right hand side is bounded by the assumptions on the initial data and the L^∞-bounds proved in Lemmas 3.2 and 3.3. This implies (3.5). Therefore the right hand side of (3.4) is contained in $L^1((0,T)\times\Omega)$. Using (3.5) for the third term on the left hand side, we conclude that it is contained in $L^1([0,T];H^{-1}(\Omega))$. The second term on the left hand side is contained in $L^\infty([0,T];W^{-1,2}(\Omega))$. Hence $\partial_t f_\varepsilon = g_\varepsilon + k_\varepsilon$ with $g_\varepsilon \in L^1([0,T]\times\Omega)$ and $k_\varepsilon \in L^1([0,T];W^{-1,2}(\Omega))$ and in particular, $\partial_t f_\varepsilon \in L^1([0,T];W^{-1,1^*}(\Omega))$ by the Sobolev embedding ($1^* = d/(d-1)$). □

REMARK 3.5. The preceeding lemma implies that the time derivative of the approximation of the pressure $\partial_t p_\varepsilon = \partial_t |n_h|^\gamma = g_\varepsilon + k_\varepsilon$ where g_ε is uniformly bounded in $L^1([0,T]\times\Omega)$ and k_ε in $L^1([0,T];H^{-1}(\Omega))$. Hence $\partial_t W_\varepsilon = U_\varepsilon + V_\varepsilon$ where $U_\varepsilon \in L^1([0,T];H^1(\Omega))$ solves $-\mu\Delta U_\varepsilon + U_\varepsilon = k_\varepsilon$ and $V_\varepsilon \in L^1([0,T];W^{1,r}(\Omega))$, $1 \le r < 1^*$ solves $-\mu\Delta V_\varepsilon + V_\varepsilon = g_\varepsilon$ (see [**13**, Thm. 6.1] for a proof of the second statement). Hence $\partial_t W_\varepsilon \in L^1([0,T];W^{1,r}(\Omega))$ for any $1 \le r < 1^*$.

3.3. Passing to the limit $\varepsilon \to 0$. The estimates of the previous (sub)sections allow us to pass to the limit $\varepsilon \to 0$ in a subsequence, still denoted ε, and conclude the existence of limit functions

$$\begin{aligned} n_\varepsilon &\rightharpoonup n \ge 0, \quad \text{in } L^q([0,T]\times\Omega),\ 1 \le q < \infty,\\ p_\varepsilon &\rightharpoonup \overline{p} \ge 0, \quad \text{in } L^q([0,T]\times\Omega),\ 1 \le q < \infty, \end{aligned}$$

where $p_\varepsilon := n_\varepsilon^\gamma$ and $0 \le n, \overline{p} \in L^\infty([0,T]\times\Omega)$. Using Aubin-Lions' lemma for W_ε and ∇W_ε, we obtain strong convergence of a subsequence in $L^q([0,T]\times\Omega)$ for any $q \in [0,\infty)$ to limit functions $W, \nabla W \in L^q([0,T]\times\Omega)$. Moreover, from the estimates in Lemma 3.3 we obtain that $W \in L^\infty([0,T]\times\Omega) \cap L^\infty([0,T];W^{2,q}(\Omega))$. Hence we have that $(n, W, \overline{p})$ satisfy for any $\varphi, \psi \in C_0^1([0,T)\times\Omega)$,

$$\begin{aligned} \int_0^T\int_\Omega n\varphi_t - n\nabla W\cdot\nabla\varphi\,dxdt + \int_\Omega n_0\,\varphi(0,x)dx &= -\int_0^T\int_\Omega \overline{n\boldsymbol{G}(p)}\varphi\,dxdt\\ \int_0^T\int_\Omega W\psi + \mu\nabla W\cdot\nabla\psi\,dxdt &= \int_0^T\int_\Omega \overline{p}\,\psi\,dxdt \end{aligned} \tag{3.6}$$

where $\overline{n\boldsymbol{G}(p)}$ is the weak limit of $n_\varepsilon \boldsymbol{G}(p_\varepsilon)$. To conclude that the limit (n, W, p) is a weak solution of (1.6), we need to show that n_ε converges strongly and therefore in the limit $\overline{p} = p := n^\gamma$ and $\overline{n\boldsymbol{G}(p)} = n\boldsymbol{G}(p)$. For this purpose, we combine a compensated compactness property (Lemma 3.7) with a monotonicity argument. We will also make use of the following lemma which was proved in a more general version in [**38, 101**]:

LEMMA 3.6. *Let $n, f \in L^\infty([0,T]\times\Omega)$ and $\boldsymbol{u} \in L^\infty([0,T];H^1(\Omega))$ with $\operatorname{div}\boldsymbol{u} \in L^\infty([0,T]\times\Omega)$ satisfy*

$$n_t - \operatorname{div}(\boldsymbol{u}n) = f, \tag{3.7}$$

in the sense of distributions. Then for all continuously differentiable functions $b \in C^1(\mathbb{R})$,

$$b(n)_t - \operatorname{div}(\boldsymbol{u} b(n)) = b'(n)f + [b'(n)n - b(n)] \operatorname{div} \boldsymbol{u}, \tag{3.8}$$

in the sense of distributions.

PROOF. We let $0 \leq \psi \in C_0^\infty(\mathbb{R}^{d+1})$ be a smooth, radially symmetric mollifier, i.e. $\psi(x) = \psi(-x)$ and $\int_{\mathbb{R}^{d+1}} \psi(x)dx$, with $\operatorname{supp}(\psi) \subset B_1(0)$ and denote for $\delta > 0$, $\psi_\delta(x) := \delta^{-(d+1)}\psi(x/\delta)$. Then we choose as a test function in (3.7) $\psi_\delta(s,y)\varphi(t+s,x+y)$, with φ is compactly supported in $(\delta, T-\delta) \times \Omega^\delta$ where Ω^δ includes all the points x in Ω which have distance $d(x, \partial\Omega) > \delta$ and do a change of variables:

$$\int_0^T \int_\Omega n(t-s,x-y)\psi_\delta(s,y)\partial_t\varphi(t,x) - n(t-s,x-y)\boldsymbol{u}(t,x)\psi_\delta(s,y)\cdot\nabla\varphi(t,x)\,dxdt \\ = -\int_0^T \int_\Omega f(t-s,x-y)\psi_\delta(s,y)\varphi(t,x)\,dxdt.$$

Integrating in (s,y), this becomes

$$\int_0^T \int_\Omega (n*\psi_\delta)(t,x)\partial_t\varphi(t,x) - (n\boldsymbol{u})*\psi_\delta(t,x)\cdot\nabla\varphi(t,x)\,dxdt \\ = -\int_0^T \int_\Omega (f*\psi_\delta)(t,x)\varphi(t,x)\,dxdt.$$

We define $n_\delta := n*\psi_\delta$ and $f_\delta := f*\psi_\delta$ and choose as a test function $\varphi := b'(n_\delta)\phi$ for a smooth ϕ compactly supported in $(\delta, T-\delta)\times\Omega^\delta$ (which is possible since n_δ is smooth and bounded thanks to the convolution.). Then we can rewrite the last identity using chain rule as

$$\int_0^T \int_\Omega b(n_\delta)\partial_t\phi - b(n_\delta)\boldsymbol{u}\cdot\nabla\phi\,dxdt \\ = -\int_0^T \int_\Omega (b'(n_\delta)f_\delta + [b'(n_\delta)n_\delta - b(n_\delta)]\operatorname{div}\boldsymbol{u} + b'(n_\delta)r_\delta)\,\phi\,dxdt.$$

where $r_\delta := \operatorname{div}((n\boldsymbol{u})*\psi_\delta) - \operatorname{div}(n_\delta\boldsymbol{u})$. By [**92**, Lemma 2.3], we have that $r_\delta \to 0$ in $L^2_{\text{loc}}((0,T)\times\Omega)$ and thanks to the properties of the convolution that $b(n_\delta) \to b(n)$ almost everywhere as well as $f_\delta \to f$ a.e. when $\delta \to 0$. Thus we obtain that in the limit $\delta \to 0$, n satisfies

$$\int_0^T \int_\Omega b(n)\partial_t\phi - b(n)\boldsymbol{u}\cdot\nabla\phi\,dxdt = -\int_0^T \int_\Omega (b'(n)f + [b'(n)n - b(n)]\operatorname{div}\boldsymbol{u})\,\phi\,dxdt.$$

which is exactly (3.8) in the sense of distributions. □

Applying Lemma 3.6 for the weak limit n in (3.6) with $b(n) = n^2$, we obtain that n satisfies

$$\int_0^T \int_\Omega n^2\varphi_t - n^2\nabla W\cdot\nabla\varphi\,dxdt = -\int_0^T \int_\Omega (2n\overline{n\boldsymbol{G}(p)} + n^2\Delta W)\varphi\,dxdt \tag{3.9}$$

for any test functions $\varphi \in C_0^1((0,T)\times\Omega)$. On the other hand, from (3.4) for $b(n)=n^2$ we obtain after integrating in space and time

$$\int_\Omega n_\varepsilon^2(\tau)\,dx - \int_\Omega n_\varepsilon^2(0)\,dx \le \int_0^\tau \int_\Omega n_\varepsilon^2 \Delta W_\varepsilon + 2n_\varepsilon^2 \boldsymbol{G}(p_\varepsilon)\,dxdt$$

Passing to the limit $\varepsilon \to 0$ in this inequality, we have

$$\int_\Omega \overline{n^2}(\tau)\,dx - \int_\Omega n_0^2\,dx \le \int_0^\tau \int_\Omega \overline{n^2\Delta W} + 2\overline{n^2\boldsymbol{G}(p)}\,dxdt, \tag{3.10}$$

where $\overline{n^2}$ denotes the weak limit of n_ε^2 and $\overline{n^2\Delta W}$ and $\overline{n^2\boldsymbol{G}(p)}$ are the weak limits of $n_\varepsilon^2\Delta W_\varepsilon$ and $n_\varepsilon^2\boldsymbol{G}(p_\varepsilon)$ respectively. Letting $\tau \to 0$ in this inequality, we obtain, thanks to the boundedness of the integrand on the right hand side,

$$\int_\Omega \overline{n^2}(0)\,dx - \int_\Omega n_0^2\,dx \le 0.$$

On the other hand, since $b(n)=n^2$ is convex, we have $\overline{n^2} \ge n^2$ and hence $\overline{n^2}(0,x) = n_0^2(x)$.

We now choose smooth test functions φ_ϵ approximating $\varphi(t,x) = \mathbf{1}_{[0,\tau]}(t)$, where $\tau \in (0,T]$, in inequality (3.9) and then pass to the limit in the approximation to obtain the inequality

$$\int_\Omega n^2(\tau)\,dx - \int_\Omega n_0^2\,dx = \int_0^\tau \int_\Omega (2n\overline{n\boldsymbol{G}(p)} + n^2\Delta W)\,dxdt \tag{3.11}$$

Subtracting (3.11) from (3.10), we have

$$\begin{aligned}\int_\Omega \left(\overline{n^2} - n^2\right)(\tau)dx \\ \le \int_0^\tau\int_\Omega \left(2\overline{n^2\boldsymbol{G}(p)} - 2n\overline{n\boldsymbol{G}(p)} + \Delta W\left(\overline{n^2}-n^2\right) + \overline{n^2\Delta W} - \overline{n^2}\Delta W\right) dx\,dt.\end{aligned} \tag{3.12}$$

Now using the explicit expression of $\boldsymbol{G}$, (1.4), the first term on the right hand side can be estimated as follows:

$$\begin{aligned}&\int_0^\tau\int_\Omega \left(2\overline{n^2\boldsymbol{G}(p)} - 2n\overline{n\boldsymbol{G}(p)}\right) dx\,dt \\ &\quad= 2\int_0^\tau\int_\Omega \alpha\left(\overline{n^2}-n^2\right) - \beta\left(\overline{n^{2+\gamma\theta}} - n\overline{n^{1+\gamma\theta}}\right) dx\,dt \\ &\quad\le 2\int_0^\tau\int_\Omega \alpha\left(\overline{n^2}-n^2\right) - \beta\left(\overline{n^{2+\gamma\theta}} - \overline{n^{2+\gamma\theta}}\right) dx\,dt \\ &\quad\le 2\alpha\int_0^\tau\int_\Omega \left(\overline{n^2}-n^2\right) dx\,dt\end{aligned} \tag{3.13}$$

where we have used [**101**, Lemma 3.35], which implies $n\overline{n^{1+\gamma\theta}} \le \overline{n^{2+\gamma\theta}}$, for the first inequality. To estimate the second term on the right hand side, we use that ΔW is bounded

thanks to Lemma 3.3 and that $\overline{n^2} \geq n^2$ by the convexity of $f(x) = x^2$. Hence

$$\int_0^\tau \int_\Omega \Delta W \left(\overline{n^2} - n^2\right) dxdt \leq \frac{P_M}{\mu} \int_0^\tau \int_\Omega \left(\overline{n^2} - n^2\right) dxdt. \tag{3.14}$$

For the last term, we use the following lemma,

LEMMA 3.7. *The weak limits $(n, W, \overline{p})$ of the sequences $\{(n_\varepsilon, W_\varepsilon, p_\varepsilon)\}_{\varepsilon>0}$ satisfy for smooth functions $S : \mathbb{R} \to \mathbb{R}$,*

$$\int_\Omega \left(\overline{S(n)\Delta W} - \overline{S(n)}\Delta W\right) dx = \frac{1}{\mu} \int_\Omega \left(\overline{p}\ \overline{S(n)} - \overline{pS(n)}\right) dx \tag{3.15}$$

where $\overline{S(n)\Delta W}$, $\overline{S(n)}$, $\overline{pS(n)}$ are the weak limits of $S(n_\varepsilon)\Delta W_\varepsilon$, $S(n_\varepsilon)$ and $p_\varepsilon S(n_\varepsilon)$ respectively.

Applying this lemma to the second term in (3.12) with $S(n) = n^2$, we can estimate it by

$$\begin{aligned}\int_0^\tau \int_\Omega \left(\overline{n^2\Delta W} - \overline{n^2}\Delta W\right) dx &= \frac{1}{\mu} \int_\Omega \left(\overline{p}\ \overline{n^2} - \overline{pn^2}\right) dxdt \\ &= \frac{1}{\mu} \int_\Omega \left(\overline{n^\gamma}\ \overline{n^2} - \overline{n^{2+\gamma}}\right) dxdt \\ &\leq 0,\end{aligned}$$

using that $\overline{n^\gamma}\ \overline{n^2} \leq \overline{n^{2+\gamma}}$ (cf. [**101**]). Thus,

$$\int_\Omega \left(\overline{n^2} - n^2\right)(\tau) dx \leq \left(2\alpha + \frac{P_M}{\mu}\right) \int_0^\tau \int_\Omega \left(\overline{n^2} - n^2\right) dx\, dt.$$

Hence Grönwall's inequality implies

$$\int_\Omega \left(\overline{n^2} - n^2\right)(\tau) dx \leq 0$$

By convexity of the function $f(x) = x^2$ we also have $n^2 \leq \overline{n^2}$ almost everywhere and so

$$\overline{n^2}(t, x) = n^2(t, x)$$

almost everywhere in $(0, T) \times \Omega$. Therefore we conclude that the functions n_ε converge strongly to n almost everywhere and in particular also $\overline{p} = n^\gamma$ which means that the limit $(n, W, \overline{p})$ is a weak solution of the equations (1.6).

PROOF OF LEMMA 3.7. We multiply the equation for W_ε by $S(n_\varepsilon)$ and integrate over Ω,

$$\int_\Omega \mu\Delta W_\varepsilon\, S(n_\varepsilon) - W_\varepsilon S(n_\varepsilon)\, dx = -\int_\Omega p_\varepsilon S(n_\varepsilon)\, dx.$$

Passing to the limit $\varepsilon \to 0$, we obtain

$$\int_\Omega \mu\overline{\Delta W\, S(n)} - W\overline{S(n)}\, dx = -\int_\Omega \overline{pS(n)}\, dx. \tag{3.16}$$

On the other hand, using the smooth function $S(n_\varepsilon)$ as a test function in the weak formulation of the limit equation

$$-\mu\Delta W + W = \overline{p},$$

and passing to the limit $\varepsilon \to 0$, we obtain

$$\int_\Omega \mu\Delta W\overline{S(n)} - W\overline{S(n)}\,dx = -\int_\Omega \overline{p}\;\overline{S(n)}\,dx.$$

Combining the last identity with (3.16), we obtain (3.15). □

4. Global existence via a numerical approximation

We consider the problem in two space dimensions in a rectangular domain, for simplicity we use $\Omega = [0,1]^2$, the generalization to other rectangular domains as well as three space dimensions is straightforward but more cumbersome in terms of notation, for this reason we restrict ourself to a square two dimensional domain here. For simplicity, we will also assume $a = 1$ in the Brinkman law in (1.6). We let $h > 0$ the mesh width, and Δt the time step size. We will determine the necessary ratio between h and Δt later on. For $i, j = 1, \dots, N_x$, where $N_x = 1/h$, h chosen such that N_x is an integer, we denote grid cells $\mathcal{C}_{ij} := ((i-1)h, ih] \times ((j-1)h, jh]$ with cell midpoints $x_{i,j} = ((i-1/2)h, (j-1/2)h)$. In addition, we denote $t^m = m\Delta t$, $m = 0, \dots N_T$, where $N_T = T/\Delta t$ for some final time $T > 0$. The approximation of a function f at grid point $x_{i,j}$ and time t^m will be denoted $f^m_{i,j}$. We also introduce the finite differences,

$$D_1^\pm f_{ij} = \pm\frac{f_{i\pm1,j} - f_{i,j}}{h}, \quad D_2^\pm f_{ij} = \pm\frac{f_{i,j\pm1} - f_{i,j}}{h}, \quad D_t^\pm f^m = \pm\frac{f^{m\pm1} - f^m}{\Delta t}.$$

and define the discrete Laplacian, divergence and gradient operators based on these,

$$\nabla_h^\pm := (D_1^\pm, D_2^\pm)^t, \quad \operatorname{div}_h^\pm f_{i,j} = D_1^\pm f^{(1)}_{i,j} + D_2^\pm f^{(2)}_{i,j}, \quad \Delta_h := \operatorname{div}_h^\pm \nabla_h^\mp.$$

For ease of notation, we also let $u_{i+1/2,j}$ and $v_{i,j+1/2}$ denote the discrete velocities in the transport equation, specifically, given $W_{i,j}$, we let

$$u_{i+1/2,j} := D_1^+ W_{i,j}, \quad v_{i,j+1/2} := D_2^+ W_{i,j}. \tag{4.1}$$

4.1. An explicit finite difference scheme. Given $(n^m_{i,j}, W^m_{i,j})$ at time step m, we define the quantities $(n^{m+1}_{i,j}, W^{m+1}_{i,j})$ at the next time step by

$$-\mu\Delta_h W^m_{i,j} + W^m_{i,j} = p^m_{i,j}, \tag{4.2a}$$

$$p^m_{i,j} := |n^m_{i,j}|^\gamma, \tag{4.2b}$$

$$D_t^+ n^m_{i,j} + D_1^- F^{(1)}_{i+1/2,j}(u^m, n^m) + D_2^- F^{(2)}_{i,j+1/2}(v^m, n^m) = n^m_{i,j}\boldsymbol{G}(p^m_{i,j}), \tag{4.2c}$$

where $p_{i,j} = (n_{i,j})^\gamma$ and the fluxes $F^{(j)}$, $j = 1, 2$ are defined by

$$\begin{aligned} F^{(1)}_{i+1/2,j}(u^m, n^m) &= -u^m_{i+1/2,j}\frac{n^m_{i,j} + n^m_{i+1,j}}{2} - \frac{h}{2}|u_{i+1/2,j}|D_1^+ n^m_{i,j} \\ F^{(2)}_{i,j+1/2}(v^m, n^m) &= -v^m_{i,j+1/2}\frac{n^m_{i,j} + n^m_{i,j+1}}{2} - \frac{h}{2}|v_{i,j+1/2}|D_2^+ n^m_{i,j}. \end{aligned} \tag{4.3}$$

We use homogeneous Neumann or periodic boundary conditions for both variables:

$$\begin{aligned} n^m_{0,j} &= n^m_{1,j}, \qquad n^m_{N_x+1,j} = n^m_{N_x,j}, && j = 1,\dots,N_x,\\ n^m_{i,0} &= n^m_{i,1}, \qquad n^m_{i,N_x+1} = n^m_{i,N_x}, && i = 1,\dots,N_x,\\ W^m_{0,j} &= W^m_{1,j}, \qquad W^m_{N_x+1,j} = W^m_{N_x,j}, && j = 1,\dots,N_x,\\ W^m_{i,0} &= W^m_{i,1}, \qquad W^m_{i,N_x+1} = W^m_{i,N_x}, && i = 1,\dots,N_x. \end{aligned}$$

The initial condition we approximate taking averages over the cells,

$$n^0_{i,j} = \frac{1}{|\mathcal{C}_{ij}|}\int_{\mathcal{C}_{ij}} n_0(x)\,dx, \quad p^0_{i,j} = |n^0_{i,j}|^\gamma, \quad i,j = 1,\dots,N_x.$$

4.2. Estimates on approximations. In the following, we will prove estimates on the discrete quantities $(n^m_{i,j}, W^m_{i,j})$ obtained using the scheme (4.1)–(4.3). We therefore define the piecewise constant functions

$$f_h(t,x) = \sum_{m=0}^{N_T}\sum_{i,j=1}^{N_x} f^m_{i,j}\,\mathbf{1}_{\mathcal{C}_{ij}}(x)\mathbf{1}_{[t^m,t^{m+1})}(t), \quad (t,x)\in[0,T]\times\Omega, \tag{4.4}$$

where $f \in \{n, W, p\}$. We first prove that n_h stays nonnegative and uniformly bounded from above.

LEMMA 4.1. *If $0 \le n^0_{i,j} \le n_\infty := P_M^{1/\gamma} < \infty$ uniformly in $h > 0$ and the timestep Δt satisfies the CFL condition*

$$\Delta t \le \min\left\{\frac{h}{8\max_{ij}|\nabla_h W^m_{i,j}| + h\boldsymbol{G}^\infty}, \frac{\mu}{4\gamma\overline{n}_\infty^\gamma}\right\} \tag{4.5}$$

(where $\boldsymbol{G}^\infty := \max_{s\in\mathbb{R}^+}\boldsymbol{G}(s)$), then for any $t > 0$, the functions $n_h(t,\cdot)$ are uniformly (in $h > 0$) bounded and nonnegative, specifically, defining $\overline{n}_\infty = n_\infty + 4\Delta t\sup_{s\ge0}\left(s^{1/\gamma}\boldsymbol{G}(s)\right)$, we have for all $m \ge 0$,

$$0 \le \min_{i,j} n^m_{i,j} \le \max_{i,j} n^m_{i,j} \le \overline{n}_\infty.$$

PROOF. The proof goes by induction on the timestep m. Clearly, by the assumptions, we have $0 \le n^0_{i,j} \le \overline{n}_\infty$. For the induction step we therefore assume that this holds for timestep $m > 0$ and show that it implies the nonnegativity and boundedness at timestep $m+1$.

We first show that the $W^m_{i,j}$ are bounded in terms of the $p^m_{i,j}$. To do so, let us assume it has a local maximum $W^m_{\hat\imath,\hat\jmath}$ in a cell $\mathcal{C}_{\hat\imath\hat\jmath}$, for some $\hat\imath,\hat\jmath \in \{1,\dots,N_x\}$. Then

$$D^+_k W^m_{\hat\imath,\hat\jmath} \le 0, \quad -D^-_k W^m_{\hat\imath,\hat\jmath} \le 0, \quad k = 1,2,$$

(if $\hat\imath$ or $\hat\jmath \in \{1, N_x\}$, then because of the Neumann boundary conditions, the forward/backward difference in direction of the boundary is zero and thus the previous inequality is true as well). Hence

$$\Delta_h W^m_{\hat\imath,\hat\jmath} = \frac{1}{h}\sum_{k=1}^{2}\left(D^+_k W^m_{\hat\imath,\hat\jmath} - D^-_k W^m_{\hat\imath,\hat\jmath}\right) \le 0.$$

Therefore,

$$W^m_{\hat{i},\hat{j}} = p^m_{\hat{i},\hat{j}} + \frac{1}{\mu}\Delta_h W^m_{\hat{i},\hat{j}} \le p^m_{\hat{i},\hat{j}} \le \max_{i,j} |n^m_{i,j}|^\gamma.$$

Similarly, at a local minimum $W^m_{\hat{i},\hat{j}}$ of W_h, we have

$$D^+_k W^m_{\hat{i},\hat{j}} \ge 0, \quad -D^-_k W^m_{\hat{i},\hat{j}} \ge 0, \quad k = 1, 2,$$

and hence

$$\Delta_h W^m_{\hat{i},\hat{j}} = \frac{1}{h}\sum_{k=1}^{2} \left(D^+_k W^m_{\hat{i},\hat{j}} - D^-_k W^m_{\hat{i},\hat{j}}\right) \ge 0,$$

which implies

$$W^m_{\hat{i},\hat{j}} = p^m_{\hat{i},\hat{j}} + \frac{1}{\mu}\Delta_h W^m_{\hat{i},\hat{j}} \ge p^m_{\hat{i},\hat{j}} \ge \min_{i,j} |n^m_{i,j}|^\gamma \ge 0.$$

Thus,

$$0 \le W_h \le \max_{i,j} |n^m_{i,j}|^\gamma. \tag{4.6}$$

Now we rewrite the scheme (4.2c) as

$$n^{m+1}_{i,j} = \left(\alpha^{(1),m}_{i,j} + \alpha^{(2),m}_{i,j}\right) n^m_{i,j} + \beta^m_{i,j} n^m_{i+1,j} + \zeta^m_{i,j} n^m_{i-1,j} + \eta^m_{i,j} n^m_{i,j+1} + \theta^m_{i,j} n^m_{i,j-1} \tag{4.7}$$

where

$$\begin{aligned}
\alpha^{(1),m}_{i,j} &= 1 - \frac{\Delta t}{2h}\big[(|u^m_{i+1/2,j}| + u^m_{i+1/2,j}) + (|u^m_{i-1/2,j}| - u^m_{i-1/2,j}) \\
&\qquad + (|v^m_{i,j+1/2}| + v^m_{i,j+1/2}) + (|v^m_{i,j-1/2}| - v^m_{i,j-1/2})\big] \\
\alpha^{(2),m}_{i,j} &= \Delta t\, \boldsymbol{G}(p^m_{i,j}) + \frac{\Delta t}{h}\left[u^m_{i+1/2,j} - u^m_{i-1/2,j} + v^m_{i,j+1/2} - v^m_{i,j-1/2}\right] \\
\beta^m_{i,j} &= \frac{\Delta t}{2h}\left(u^m_{i+1/2,j} + |u^m_{i+1/2,j}|\right) \\
\zeta^m_{i,j} &= \frac{\Delta t}{2h}\left(|u^m_{i-1/2,j}| - u^m_{i-1/2,j}\right) \\
\eta^m_{i,j} &= \frac{\Delta t}{2h}\left(v^m_{i,j+1/2} + |v^m_{i,j+1/2}|\right) \\
\theta^m_{i,j} &= \frac{\Delta t}{2h}\left(|v^m_{i,j-1/2}| - v^m_{i,j-1/2}\right)
\end{aligned}$$

We note that $\beta^m_{i,j}, \zeta^m_{i,j}, \eta^m_{i,j}, \theta^m_{i,j} \ge 0$, and that under the CFL-condition (4.5), also $\alpha^{(1),m}_{i,j} + \alpha^{(2),m}_{i,j} \ge 0$. Hence, assuming that $n^m_{i,j} \ge 0$ for all i, j, we have

$$\begin{aligned}
n^{m+1}_{i,j} &\ge \left(\beta^m_{i,j} + \zeta^m_{i,j} + \eta^m_{i,j} + \theta^m_{i,j}\right) \min\{n^m_{i+1,j}, n^m_{i-1,j}, n^m_{i,j+1}, n^m_{i,j-1}\} \\
&\quad + \left(\alpha^{(1),m}_{i,j} + \alpha^{(2),m}_{i,j}\right) n^m_{i,j} \\
&\ge 0.
\end{aligned}$$

We proceed to showing the boundedness of n_h. Thanks to the CFL-condition (4.5), we have

$$\alpha_{i,j}^{(1),m} \geq \frac{1}{2}, \quad \beta_{i,j}^m, \zeta_{i,j}^m, \eta_{i,j}^m, \theta_{i,j}^m \leq \frac{1}{8}.$$

Moreover, $\alpha_{i,j}^{(1),m} + \beta_{i,j}^m + \zeta_{i,j}^m + \eta_{i,j}^m + \theta_{i,j}^m = 1$. Using the induction hypothesis that $n_{i,j}^m \leq \overline{n}_\infty$ for all i, j and the nonnegativity of n_h which we have just proved, we can estimate $n_{i,j}^{m+1}$:

$$\begin{aligned} n_{i,j}^{m+1} &\leq \left(\alpha_{i,j}^{(1),m} + \alpha_{i,j}^{(2),m}\right) n_{i,j}^m + \left(\beta_{i,j}^m + \zeta_{i,j}^m + \eta_{i,j}^m + \theta_{i,j}^m\right) \overline{n}_\infty \\ &\leq \left(\frac{1}{2} + \alpha_{i,j}^{(2),m}\right) n_{i,j}^m + \frac{1}{2}\overline{n}_\infty \\ &= \overline{n}_\infty - \frac{1}{2}\left(\overline{n}_\infty - n_{i,j}^m\right) + \alpha_{i,j}^{(2),m} n_{i,j}^m \end{aligned} \tag{4.8}$$

We can rewrite and bound $\alpha_{i,j}^{(2),m}$ using the equation for $W_{i,j}^m$, (4.2a),

$$\begin{aligned} \alpha_{i,j}^{(2),m} &= \Delta t \left(\boldsymbol{G}(p_{i,j}^m) + \Delta_h W_{i,j}^m\right) \\ &= \Delta t \left(\boldsymbol{G}(p_{i,j}^m) + \frac{1}{\mu}\left(W_{i,j}^m - p_{i,j}^m\right)\right) \\ &\leq \Delta t \left(\boldsymbol{G}(p_{i,j}^m) + \frac{1}{\mu}\left(\overline{n}_\infty^\gamma - |n_{i,j}^m|^\gamma\right)\right) \\ &\leq \Delta t \left(\boldsymbol{G}(p_{i,j}^m) + \frac{\gamma\, \overline{n}_\infty^{\gamma-1}}{\mu}\left(\overline{n}_\infty - n_{i,j}^m\right)\right) \\ &\leq \Delta t \boldsymbol{G}(p_{i,j}^m) + \frac{1}{4\overline{n}_\infty}(\overline{n}_\infty - n_{i,j}^m), \end{aligned}$$

where we have used (4.6) for the first inequality, that $f(a) - f(b) = f'(\widetilde{a})(a-b)$ for some intermediate value $\widetilde{a} \in [b, a]$, with $f(a) = a^\gamma$, for the second inequality and the CFL-condition for the last inequality. Now going back to (4.8) and inserting this there, we obtain,

$$\begin{aligned} n_{i,j}^{m+1} &\leq \overline{n}_\infty - \frac{1}{2}\left(\overline{n}_\infty - n_{i,j}^m\right) + \left(\Delta t \boldsymbol{G}(p_{i,j}^m) + \frac{1}{4\overline{n}_\infty}(\overline{n}_\infty - n_{i,j}^m)\right) n_{i,j}^m \\ &\leq \frac{3}{4}\overline{n}_\infty + \frac{1}{4} n_{i,j}^m + \Delta t n_{i,j}^m \boldsymbol{G}(p_{i,j}^m) \end{aligned} \tag{4.9}$$

If $n_{i,j}^m \geq n_\infty$ then $\boldsymbol{G}(p_{i,j}^m) \leq 0$ and hence the expression in (4.9) is bounded by $\overline{n}_\infty$. On the other hand, if $n_{i,j}^m \leq n_\infty$, we can bound it by

$$\begin{aligned} n_{i,j}^{m+1} &\leq \frac{3}{4}\overline{n}_\infty + \frac{1}{4} n_{i,j}^m + \Delta t n_{i,j}^m \boldsymbol{G}(p_{i,j}^m) \\ &\leq \frac{3}{4}\overline{n}_\infty + \frac{1}{4}\left(n_\infty + 4\Delta t \sup_{s \geq 0}\left(s^{1/\gamma}\boldsymbol{G}(s)\right)\right) \\ &= \overline{n}_\infty \end{aligned}$$

where we used the definition of $\overline{n}_\infty$ for the last equality. This proves that $n_{i,j}^{m+1} \leq \overline{n}_\infty$ for all i, j if the same holds already for the $n_{i,j}^m$. □

REMARK 4.2. The estimates in the proof of the previous lemma are very coarse and therefore one can use a much larger CFL-condition than (4.5) in practice. Also note that $\overline{n}_\infty \to n_\infty$ when $\Delta t \to 0$.

4.2.1. *Estimates on the discrete potential W_h.*

LEMMA 4.3. *We have that*

$$W_h, \nabla_h W_h, \nabla_h^2 W_h \subset L^\infty([0,T]; L^2(\Omega)),$$

uniformly in $h > 0$, where $\nabla_h := \nabla_h^\pm$ and $\nabla_h^2 := \nabla_h^\mp \nabla_h^\pm$ and

$$W_h, \Delta_h W_h \subset L^\infty((0,T) \times \Omega)),$$

uniformly in $h > 0$ as well.

PROOF. To obtain the L^2-estimates, we square the equation for the potential W_h, (4.2a) and sum over all i, j,

$$\mu^2 \sum_{i,j=1}^{N_x} |\Delta_h W_{i,j}^m|^2 - 2\mu \sum_{i,j=1}^{N_x} W_{i,j}^m \Delta_h W_{i,j}^m + \sum_{i,j=1}^{N_x} |W_{i,j}^m|^2 = \sum_{i,j=1}^{N_x} |n_{i,j}^m|^{2\gamma}.$$

Using summation by parts and that W satisfies either periodic or homogeneous Neumann boundary conditions, we obtain

$$\mu^2 \sum_{i,j=1}^{N_x} |\nabla_h^2 W_{i,j}^m|^2 + 2\mu \sum_{i,j=1}^{N_x} |\nabla_h W_{i,j}^m|^2 + \sum_{i,j=1}^{N_x} |W_{i,j}^m|^2 = \sum_{i,j=1}^{N_x} |n_{i,j}^m|^{2\gamma}.$$

From the previous estimates, we know that $n_h \in L^\infty([0,T] \times \Omega)$ uniformly in $h > 0$ and therefore also uniformly bounded in any other L^p-space, which implies together with the above identity, that $W_h, \nabla_h W_h, \nabla_h^2 W_h \in L^2([0,T] \times \Omega)$. That W_h is uniformly bounded follows from (4.6) and the uniform bound on n_h which was proved in the previous Lemma 4.1.

Using this and the uniform boundedness of the pressure, we conclude by (4.2a) that also $\Delta_h W_h$ is uniformly bounded. □

REMARK 4.4. Using the discrete Gagliardo-Nirenberg-Sobolev inequality [**14**, Thm. 3.4], we obtain that $\nabla_h W_h \in L^\infty([0,T]; L^q(\Omega))$ for $1 \leq q < q^* = 2d/(d-2)$.

4.3. Discrete entropy inequalities for n_h. To prove strong convergence of the approximating sequence $\{(n_h, W_h)\}_{h>0}$, it will be useful to derive entropy inequalities for n_h. To this end, the following lemma will be useful:

LEMMA 4.5. *Let $f : \mathbb{R} \to \mathbb{R}$ be a smooth convex function and assume that Δt satisfies the CFL-condition*

$$\Delta t \leq \min\left\{ \frac{h}{16 \max_{i,j} |\nabla_h W_{i,j}^m|}, \frac{h}{8 \max_{ij} |\nabla_h W_{i,j}^m| + h \boldsymbol{G}^\infty}, \frac{\mu}{4\gamma \overline{n}_\infty^\gamma} \right\} \tag{4.10}$$

Denote $f^m_{i,j} := f(n^m_{i,j})$ *and* f_h *a piecewise constant interpolation of it as in* (4.4). *Then* $f^m_{i,j}$ *satisfies the following identity*

$$D_t f^m_{i,j} = \frac{1}{2} D_1^- \left(u^m_{i+1/2,j} \left(f^m_{i,j} + f^m_{i+1,j} \right) \right) + \frac{1}{2} D_2^- \left(v^m_{i,j+1/2} \left(f^m_{i,j} + f^m_{i,j+1} \right) \right) \tag{4.11}$$

$$+ \frac{h}{4} D_1^- \left[f'(n^m_{i,j}) |u^m_{i+1/2,j}| D_1^+ n^m_{i,j} \right] + \frac{h}{4} D_2^- \left[f'(n^m_{i,j}) |v^m_{i,j+1/2}| D_2^+ n^m_{i,j} \right] \tag{4.12}$$

$$+ \frac{h}{4} D_1^+ \left[f'(n^m_{i,j}) |u^m_{i-1/2,j}| D_1^- n^m_{i,j} \right] + \frac{h}{4} D_2^+ \left[f'(n^m_{i,j}) |v^m_{i,j-1/2}| D_2^- n^m_{i,j} \right] \tag{4.13}$$

$$- \frac{h^2}{4} D_1^- \left[f''(\widetilde{n}^m_{i+1/2,j}) u^m_{i+1/2,j} |D_1^+ n^m_{i,j}|^2 \right] \tag{4.14}$$

$$- \frac{h^2}{4} D_2^- \left[f''(\widetilde{n}^m_{i,j+1/2}) v^m_{i,j+1/2} |D_2^+ n^m_{i,j}|^2 \right] \tag{4.15}$$

$$- \frac{h}{4} f''(\widehat{n}^m_{i-1/2,j}) |u^m_{i-1/2,j}| |D_1^- n^m_{i,j}|^2 - \frac{h}{4} f''(\widehat{n}^m_{i,j-1/2}) |v^m_{i,j-1/2}| |D_2^- n^m_{i,j}|^2 \tag{4.16}$$

$$- \frac{h}{4} f''(\widehat{n}^m_{i+1/2,j}) |u^m_{i+1/2,j}| |D_1^+ n^m_{i,j}|^2 - \frac{h}{4} f''(\widehat{n}^m_{i,j+1/2}) |v^m_{i,j+1/2}| |D_2^+ n^m_{i,j}|^2 \tag{4.17}$$

$$+ (f'(n^m_{i,j}) n^m_{i,j} - f^m_{i,j}) \Delta_h W^m_{i,j} + f'(n^m_{i,j}) n^m_{i,j} \boldsymbol{G}(p^m_{i,j}) \tag{4.18}$$

$$+ \frac{\Delta t}{2} f''(\widetilde{n}^{m+1/2}_{i,j}) |D_t^+ n^m_{i,j}|^2, \tag{4.19}$$

where the intermediate values satisfy

$$\begin{aligned}
\widetilde{n}^m_{i\pm1/2,j}, \widehat{n}^m_{i\pm1/2,j} &\in [\min\{n^m_{i,j}, n^m_{i\pm1,j}\}, \max\{n^m_{i,j}, n^m_{i\pm1,j}\}], \\
\widetilde{n}^m_{i,j\pm1/2}, \widehat{n}^m_{i,j\pm1/2} &\in [\min\{n^m_{i,j}, n^m_{i,j\pm1}\}, \max\{n^m_{i,j}, n^m_{i,j\pm1}\}], \\
\widetilde{n}^{m+1/2}_{i,j} &\in [\min\{n^m_{i,j}, n^{m+1}_{i,j}\}, \max\{n^m_{i,j}, n^{m+1}_{i,j}\}],
\end{aligned}$$

and where the term (4.18) *is uniformly bounded and the terms* (4.16) – (4.17) *and* (4.19) *satisfy*

$$\begin{aligned}
\frac{h^{d+1} \Delta t}{2} \sum_{m=0}^{N_T} \sum_{i,j} f''(\widehat{n}^m_{i+1/2,j}) |u^m_{i+1/2,j}| |D_1^+ n^m_{i,j}|^2 &\le C, \\
\frac{h^{d+1} \Delta t}{2} \sum_{m=0}^{N_T} \sum_{i,j} f''(\widehat{n}^m_{i,j+1/2}) |v^m_{i,j+1/2}| |D_2^+ n^m_{i,j}|^2 &\le C, \\
\frac{h^{d} \Delta t^2}{2} \sum_{m=0}^{N_T} \sum_{i,j} f''(\widetilde{n}^{m+1/2}_{i,j}) |D_t^+ n^m_{i,j}|^2 &\le C,
\end{aligned} \tag{4.20}$$

In particular, this implies that the piecewise constant interpolation $D_t^+ f_h$ *is of the form* $D_t^+ f_h = g_h + k_h$ *where* $g_h \in L^1([0,T] \times \Omega)$ *and* $k_h \in L^\infty([0,T]; W^{-1,q}(\Omega))$ *for any* $1 \le q < \infty$ *if* $d = 2$ *and for* $1 \le q \le q^* = 2d/(d-2)$ *if* $d > 2$, *uniformly in* $h > 0$.

PROOF. We first rewrite the scheme for $n_{i,j}^m$ as

$$\begin{aligned} D_t^+ n_{i,j}^m &= \frac{1}{2} u_{i+1/2,j}^m D_1^+ n_{i,j}^m + \frac{1}{2} u_{i-1/2,j}^m D_1^- n_{i,j}^m \\ &+ \frac{1}{2} v_{i,j+1/2}^m D_2^+ n_{i,j}^m + \frac{1}{2} v_{i,j-1/2}^m D_2^- n_{i,j}^m \\ &+ \frac{h}{2} D_1^- \left[|u_{i+1/2,j}^m| D_1^+ n_{i,j}^m \right] + \frac{h}{2} D_2^- \left[|v_{i,j+1/2}^m| D_2^+ n_{i,j}^m \right] \\ &+ n_{i,j}^m \Delta_h W_{i,j}^m + n_{i,j}^m \boldsymbol{G}(p_{i,j}^m). \end{aligned} \tag{4.21}$$

Then, using the Taylor expansion,

$$f(b) - f(a) = f'(a)(b-a) + f''(\widetilde{a}) \frac{(a-b)^2}{2},$$

where $\widetilde{a} \in [\min\{a,b\}, \max\{a,b\}]$, we can write

$$\begin{aligned} D_t^+ f_{i,j}^m &= f'(n_{i,j}^m) D_t^+ n_{i,j}^m + \frac{\Delta t}{2} f''(\widetilde{n}_{i,j}^{m+1/2}) |D_t^+ n_{i,j}^m|^2 \\ D_1^\pm f_{i,j}^m &= f'(n_{i,j}^m) D_1^\pm n_{i,j}^m \pm \frac{h}{2} f''(\widetilde{n}_{i\pm1/2,j}^m) |D_1^\pm n_{i,j}^m|^2 \\ D_2^\pm f_{i,j}^m &= f'(n_{i,j}^m) D_2^\pm n_{i,j}^m \pm \frac{h}{2} f''(\widetilde{n}_{i,j\pm1/2}^m) |D_2^\pm n_{i,j}^m|^2 \\ D_1^\pm f'(n_{i,j}^m) &= f''(\widehat{n}_{i\pm1/2,j}^m) D_1^\pm n_{i,j}^m \\ D_2^\pm f'(n_{i,j}^m) &= f''(\widehat{n}_{i,j\pm1/2}^m) D_2^\pm n_{i,j}^m, \end{aligned}$$

where $\widetilde{n}_{i,j}^{m+1/2}, \widetilde{n}_{i\pm1/2,j}^m, \widetilde{n}_{i,j\pm1/2}^m, \widehat{n}_{i\pm1/2,j}^m$ and $\widehat{n}_{i,j\pm1/2}^m$ are intermediate values. Hence, multiplying equation (4.21) by $f'(n_{i,j}^m)$, it becomes

$$\begin{aligned} D_t^+ f_{i,j}^m &= \frac{\Delta t}{2} f''(\widetilde{n}_{i,j}^{m+1/2}) |D_t^+ n_{i,j}^m|^2 \\ &+ \frac{1}{2} u_{i+1/2,j}^m D_1^+ f_{i,j}^m - \frac{h}{4} f''(\widetilde{n}_{i+1/2,j}^m) u_{i+1/2,j}^m |D_1^+ n_{i,j}^m|^2 \\ &+ \frac{1}{2} u_{i-1/2,j}^m D_1^- f_{i,j}^m + \frac{h}{4} f''(\widetilde{n}_{i-1/2,j}^m) u_{i-1/2,j}^m |D_1^- n_{i,j}^m|^2 \\ &+ \frac{1}{2} v_{i,j+1/2}^m D_2^+ f_{i,j}^m - \frac{h}{4} f''(\widetilde{n}_{i,j+1/2}^m) v_{i,j+1/2}^m |D_2^+ n_{i,j}^m|^2 \\ &+ \frac{1}{2} v_{i,j-1/2}^m D_2^- f_{i,j}^m + \frac{h}{4} f''(\widetilde{n}_{i,j-1/2}^m) v_{i,j-1/2}^m |D_2^- n_{i,j}^m|^2 \\ &+ \frac{h}{4} D_1^- \left[f'(n_{i,j}^m) |u_{i+1/2,j}^m| D_1^+ n_{i,j}^m \right] - \frac{h}{4} f''(\widehat{n}_{i-1/2,j}^m) |u_{i-1/2,j}^m| |D_1^- n_{i,j}^m|^2 \\ &+ \frac{h}{4} D_2^- \left[f'(n_{i,j}^m) |v_{i,j+1/2}^m| D_2^+ n_{i,j}^m \right] - \frac{h}{4} f''(\widehat{n}_{i,j-1/2}^m) |v_{i,j-1/2}^m| |D_2^- n_{i,j}^m|^2 \\ &+ \frac{h}{4} D_1^+ \left[f'(n_{i,j}^m) |u_{i-1/2,j}^m| D_1^- n_{i,j}^m \right] - \frac{h}{4} f''(\widehat{n}_{i+1/2,j}^m) |u_{i+1/2,j}^m| |D_1^+ n_{i,j}^m|^2 \end{aligned}$$

$$
\begin{aligned}
&+\frac{h}{4}D_2^+\left[f'(n_{i,j}^m)|v_{i,j-1/2}^m|D_2^-n_{i,j}^m\right]-\frac{h}{4}f''(\widehat{n}_{i,j+1/2}^m)|v_{i,j+1/2}^m||D_2^+n_{i,j}^m|^2\\
&+f'(n_{i,j}^m)n_{i,j}^m\Delta_h W_{i,j}^m+f'(n_{i,j}^m)n_{i,j}^m\boldsymbol{G}(p_{i,j}^m)\\
=&\frac{\Delta t}{2}f''(\widetilde{n}_{i,j}^{m+1/2})|D_t^+n_{i,j}^m|^2\\
&+\frac{1}{2}D_1^-\left(u_{i+1/2,j}^m\left(f_{i,j}^m+f_{i+1,j}^m\right)\right)+\frac{1}{2}D_2^-\left(v_{i,j+1/2}^m\left(f_{i,j}^m+f_{i,j+1}^m\right)\right)\\
&+\frac{h}{4}D_1^-\left[f'(n_{i,j}^m)|u_{i+1/2,j}^m|D_1^+n_{i,j}^m\right]+\frac{h}{4}D_2^-\left[f'(n_{i,j}^m)|v_{i,j+1/2}^m|D_2^+n_{i,j}^m\right]\\
&+\frac{h}{4}D_1^+\left[f'(n_{i,j}^m)|u_{i-1/2,j}^m|D_1^-n_{i,j}^m\right]+\frac{h}{4}D_2^+\left[f'(n_{i,j}^m)|v_{i,j-1/2}^m|D_2^-n_{i,j}^m\right]\\
&-\frac{h^2}{4}D_1^-\left[f''(\widetilde{n}_{i+1/2,j}^m)u_{i+1/2,j}^m|D_1^+n_{i,j}^m|^2\right]\\
&-\frac{h^2}{4}D_2^-\left[f''(\widetilde{n}_{i,j+1/2}^m)v_{i,j+1/2}^m|D_2^+n_{i,j}^m|^2\right]\\
&-\frac{h}{4}f''(\widehat{n}_{i-1/2,j}^m)|u_{i-1/2,j}^m||D_1^-n_{i,j}^m|^2-\frac{h}{4}f''(\widehat{n}_{i,j-1/2}^m)|v_{i,j-1/2}^m||D_2^-n_{i,j}^m|^2\\
&-\frac{h}{4}f''(\widehat{n}_{i+1/2,j}^m)|u_{i+1/2,j}^m||D_1^+n_{i,j}^m|^2-\frac{h}{4}f''(\widehat{n}_{i,j+1/2}^m)|v_{i,j+1/2}^m||D_2^+n_{i,j}^m|^2\\
&+(f'(n_{i,j}^m)n_{i,j}^m-f_{i,j}^m)\Delta_h W_{i,j}^m+f'(n_{i,j}^m)n_{i,j}^m\boldsymbol{G}(p_{i,j}^m).
\end{aligned}
$$

which implies (4.11)–(4.19). In particular, for $f(x)=x^2$, this becomes

$$
\begin{aligned}
D_t^+f_{i,j}^m=&\Delta t|D_t^+n_{i,j}^m|^2\\
&+\frac{1}{2}D_1^+\left(u_{i-1/2,j}^m\left(f_{i,j}^m+f_{i-1,j}\right)\right)+\frac{1}{2}D_2^+\left(v_{i,j-1/2}^m\left(f_{i,j}^m+f_{i,j-1}\right)\right)\\
&-\frac{h^2}{2}D_1^-\left[u_{i+1/2,j}^m|D_1^+n_{i,j}^m|^2\right]-\frac{h^2}{2}D_2^-\left[v_{i,j+1/2}^m|D_2^+n_{i,j}^m|^2\right]\\
&+\frac{h}{2}D_1^-\left[n_{i,j}^m|u_{i+1/2,j}^m|D_1^+n_{i,j}^m\right]+\frac{h}{2}D_2^-\left[n_{i,j}^m|v_{i,j+1/2}^m|D_2^+n_{i,j}^m\right]\\
&+\frac{h}{2}D_1^+\left[n_{i,j}^m|u_{i-1/2,j}^m|D_1^-n_{i,j}^m\right]+\frac{h}{2}D_2^+\left[n_{i,j}^m|v_{i,j-1/2}^m|D_2^-n_{i,j}^m\right]\\
&-\frac{h}{2}|u_{i-1/2,j}^m||D_1^-n_{i,j}^m|^2-\frac{h}{2}|v_{i,j-1/2}^m||D_2^-n_{i,j}^m|^2\\
&-\frac{h}{2}|u_{i+1/2,j}^m||D_1^+n_{i,j}^m|^2-\frac{h}{2}|v_{i,j+1/2}^m||D_2^+n_{i,j}^m|^2\\
&+f_{i,j}^m\Delta_h W_{i,j}^m+2f_{i,j}^m\boldsymbol{G}(p_{i,j}^m),
\end{aligned}
\tag{4.22}
$$

We estimate the first term on the right hand side of the inequality inserting (4.21),

$$
|D_t^+n_{i,j}^m|^2\le 2\left|\frac{1}{2}u_{i+1/2,j}^mD_1^+n_{i,j}^m+\frac{1}{2}u_{i-1/2,j}^mD_1^-n_{i,j}^m+\frac{1}{2}v_{i,j+1/2}^mD_2^+n_{i,j}^m\right.
$$

$$
\begin{aligned}
&\quad + \frac{1}{2} v^m_{i,j-1/2} D_2^- n^m_{i,j} + \frac{h}{2} D_1^- \left[|u^m_{i+1/2,j}| D_1^+ n^m_{i,j} \right] + \frac{h}{2} D_2^- \left[|v^m_{i,j+1/2}| D_2^+ n^m_{i,j} \right] \Big|^2 \\
&\quad + 2 \left| n^m_{i,j} \Delta_h W^m_{i,j} + n^m_{i,j} \boldsymbol{G}(p^m_{i,j}) \right|^2 \\
&\le 4 \left| \frac{1}{2} u^m_{i+1/2,j} D_1^+ n^m_{i,j} + \frac{1}{2} u^m_{i-1/2,j} D_1^- n^m_{i,j} + \frac{h}{2} D_1^- \left[|u^m_{i+1/2,j}| D_1^+ n^m_{i,j} \right] \right|^2 \\
&\quad + 4 \left| \frac{1}{2} v^m_{i,j+1/2} D_2^+ n^m_{i,j} + \frac{1}{2} v^m_{i,j-1/2} D_2^- n^m_{i,j} + \frac{h}{2} D_2^- \left[|v^m_{i,j+1/2}| D_2^+ n^m_{i,j} \right] \right|^2 \\
&\quad + 2 \left| n^m_{i,j} \Delta_h W^m_{i,j} + n^m_{i,j} \boldsymbol{G}(p^m_{i,j}) \right|^2 \\
&\le 8 \left| u^m_{i+1/2,j} D_1^+ n^m_{i,j} \right|^2 + 8 \left| u^m_{i-1/2,j} D_1^- n^m_{i,j} \right|^2 + 8 \left| v^m_{i,j+1/2} D_2^+ n^m_{i,j} \right|^2 \\
&\quad + 8 \left| u^m_{i,j-1/2} D_2^- n^m_{i,j} \right|^2 + 2 \left| n^m_{i,j} \Delta_h W^m_{i,j} + n^m_{i,j} \boldsymbol{G}(p^m_{i,j}) \right|^2 \\
&\le 8 \max_{i,j} |\nabla_h W^m_{i,j}| \big\{ |u^m_{i+1/2,j}| \left| D_1^+ n^m_{i,j} \right|^2 + |u^m_{i-1/2,j}| \left| D_1^- n^m_{i,j} \right|^2 \\
&\qquad\qquad + |v^m_{i,j+1/2}| \left| D_2^+ n^m_{i,j} \right|^2 + |v^m_{i,j-1/2}| \left| D_2^- n^m_{i,j} \right|^2 \big\} \\
&\quad + 2 \left| n^m_{i,j} \Delta_h W^m_{i,j} + n^m_{i,j} \boldsymbol{G}(p^m_{i,j}) \right|^2
\end{aligned}
$$

Thus if we assume that Δt satisfies the CFL-condition (4.10), we have

$$
\begin{aligned}
\Delta t \sum_{i,j} |D_t^+ n^m_{i,j}|^2 &\le h \sum_{i,j} \big\{ |u^m_{i+1/2,j}| |D_1^+ n^m_{i,j}|^2 + |v^m_{i,j+1/2}| |D_2^+ n^m_{i,j}|^2 \big\} \\
&\quad + h \sum_{i,j} \left| n^m_{i,j} \Delta_h W^m_{i,j} + n^m_{i,j} \boldsymbol{G}(p^m_{i,j}) \right|^2
\end{aligned}
$$

Now summing (4.22) over all i, j, multiplying with h^d and using the latter inequality, we obtain

$$
\begin{aligned}
h^d D_t^+ \sum_{i,j} f^m_{i,j} &= -h^{d+1} \sum_{i,j} \left(|u^m_{i-1/2,j}| |D_1^- n^m_{i,j}|^2 + |v^m_{i,j-1}| |D_2^- n^m_{i,j}|^2 \right) \\
&\quad + h^d \Delta t \sum_{i,j} |D_t^+ n^m_{i,j}|^2 + h^d \sum_{i,j} f^m_{i,j} \left(\Delta_h W^m_{i,j} + 2 \boldsymbol{G}(p^m_{i,j}) \right) \\
&\le h^d \sum_{i,j} f^m_{i,j} \left(\Delta_h W^m_{i,j} + 2 \boldsymbol{G}(p^m_{i,j}) \right) \\
&\quad + h^{d+1} \sum_{i,j} |n^m_{i,j} \Delta_h W^m_{i,j} + n^m_{i,j} \boldsymbol{G}(p^m_{i,j})|^2 \\
&\le C,
\end{aligned}
$$

where $C > 0$ is a constant independent of h, thanks to the L^∞-bounds on n_h and $\Delta_h W_h$ obtained in Lemma 4.1 and 4.3. This implies that

$$h^{d+1}\Delta t \sum_{m=0}^{N_T}\sum_{i,j}\left(|u^m_{i-1/2,j}||D_1^- n^m_{i,j}|^2 + |v^m_{i,j-1/2}||D_2^- n^m_{i,j}|^2\right) \le C$$

$$h^d \Delta t^2 \sum_{m=0}^{N_T}\sum_{i,j}|D_t^+ n^m_{i,j}|^2 \le C.$$

and therefore using Hölder's inequality and the uniform L^∞-bounds on n_h, (4.20). Using summation by parts, we realize that the other terms, (4.11) – (4.15) are in $L^\infty([0,T];W^{-1,q}(\Omega))$ for $q \in [1, 2^*)$ where $2^* = 2d/(d-2)$ if $d \ge 3$ and any finite number greater than one if $d = 2$. □

Remark 4.6. The preceeding lemma implies that the forward time difference of the approximation of the pressure $D_t^+ p_h = D_t^+ |n_h|^\gamma$ is of the form $D_t^+ p_h = g_h + k_h$ where $g_h \in L^1([0,T]\times\Omega)$ and $k_h \in L^\infty([0,T];W^{-1,q}(\Omega))$ for any $1 \le q < \infty$ if $d = 2$ and for $1 \le q \le q^* = 2d/(d-2)$ if $d > 2$, uniformly in $h > 0$. Using this, we have that $D_t^+ W_h = U_h + V_h$ where U_h and V_h solve

$$-\mu\Delta_h U_h + U_h = g_h, \quad -\mu\Delta_h V_h + V_h = k_h.$$

By Lemma 6.3, we have $U_h, \nabla_h U_h \in L^1([0,T];L^q(\Omega))$ for $1 \le q \le d/(d-1)$ and by standard results, V_h, $\nabla_h V_h \in L^\infty([0,T];L^2(\Omega))$. Hence $D_t W_h, D_t\nabla_h W_h \in L^1([0,T];L^q(\Omega)) + L^\infty([0,T];L^2(\Omega))$.

Remark 4.7 (CFL-condition). The estimates from Lemma 4.3 imply that the velocity $\boldsymbol{u}_h := \nabla_h W_h \in L^\infty([0,T];L^{2^*}(\Omega))$ uniformly in $h > 0$, $2^* = 2d/(d-2)$ or any number in $[1,\infty)$ if $d = 2$, using the Sobolev embedding theorem. Using an inverse inequality, we can bound it in the $L^\infty((0,T)\times\Omega)$-norm as follows:

$$\max_{(x,t)\in(0,T)\times\Omega}|\boldsymbol{u}_h| \le Ch^{-\frac{d}{2^*}}\left(\int_\Omega |\boldsymbol{u}_h|^{2^*}dx\right)^{\frac{1}{2^*}} \le Ch^{-\frac{d}{2^*}}$$

Thus the time step size Δt is of order $\mathcal{O}(h^{1+d/2^*})$. In practice a linear CFL-condition seems to work well though.

4.4. Passing to the limit $h \to 0$. The estimates of the previous (sub)sections allow us to pass to the limit $h \to 0$ in a subsequence still denoted h,

$$n_h \rightharpoonup n \ge 0, \quad \text{in } L^q([0,T]\times\Omega),\ 1 \le q < \infty,$$
$$p_h \rightharpoonup \overline{p} \ge 0, \quad \text{in } L^q([0,T]\times\Omega),\ 1 \le q < \infty,$$

where $p_h := n_h^\gamma$ and $0 \le n, \overline{p} \in L^\infty([0,T]\times\Omega)$. Using the "discretized" Aubin-Lions lemma 6.1 for W_h and $\nabla_h W_h$, we obtain strong convergence of a subsequence in $L^q([0,T]\times\Omega)$ for any $q \in [0,\infty)$ in the case of W_h and $1 \le q \le 2^*$ in the case of $\nabla_h W_h$ ($2^* = 2d/(d-2)$ if $d \ge 3$ and any finite number greater than or equal to one if $d = 2$), to limit functions $W, \nabla W \in L^q([0,T]\times\Omega)$. Moreover, from the estimates in Lemma 4.3 we obtain that

$W \in L^\infty([0,T]\times\Omega)\cap L^\infty([0,T];H^2(\Omega))$. Hence we have that $(n,W,\overline{p})$ satisfy for any $\varphi,\psi\in C^1([0,T]\times\Omega)$,

$$\int_0^T\int_\Omega n\varphi_t - n\nabla W\cdot\nabla\varphi\,dxdt = -\int_0^T\int_\Omega \overline{n\boldsymbol{G}(p)}\varphi\,dxdt$$
$$\int_0^T\int_\Omega W\psi+\mu\nabla W\cdot\nabla\psi\,dxdt = \int_0^T\int_\Omega \overline{p}\,\psi\,dxdt$$

where $\overline{n\boldsymbol{G}(p)}$ is the weak limit of $n_h\boldsymbol{G}(p_h)$. To conclude that the limit (n,W,p) is a weak solution of (1.6), we proceed as in the previous Section 4 and show that n_h in fact converges strongly: First, we recall that the limit n satisfies (3.9).

On the other hand, from (4.22), we obtain (under the CFL-condtion (4.10))

$$\begin{aligned}
D_t^+|n_{i,j}^m|^2 \le{}& \frac{1}{2}D_1^+\left(u_{i-1/2,j}^m\left(|n_{i,j}^m|^2+|n_{i-1,j}^m|^2\right)\right)\\
&+\frac{1}{2}D_2^+\left(v_{i,j-1/2}^m\left(|n_{i,j}^m|^2+|n_{i,j-1}^m|^2\right)\right)\\
&-\frac{h^2}{2}D_1^-\left[u_{i+1/2,j}^m|D_1^+n_{i,j}^m|^2\right]-\frac{h^2}{2}D_2^-\left[v_{i,j+1/2}^m|D_2^+n_{i,j}^m|^2\right]\\
&+\frac{h}{2}D_1^-\left[n_{i,j}^m|u_{i+1/2,j}|D_1^+n_{i,j}^m\right]+\frac{h}{2}D_2^-\left[n_{i,j}^m|v_{i,j+1/2}|D_2^+n_{i,j}^m\right]\\
&+\frac{h}{2}D_1^+\left[n_{i,j}^m|u_{i-1/2,j}|D_1^-n_{i,j}^m\right]+\frac{h}{2}D_2^+\left[n_{i,j}^m|v_{i,j-1/2}|D_2^-n_{i,j}^m\right]\\
&+|n_{i,j}^m|^2\Delta_hW_{i,j}^m+2|n_{i,j}^m|^2\boldsymbol{G}(p_{i,j}^m),
\end{aligned}\tag{4.23}$$

Considering this inequality in terms of the piecewise constant functions n_h, W_h and p_h, multiplying it with a nonnegative C^1-test function φ, integrating and then passing to the limit $h\to 0$, we obtain (using the bounds (4.20), the weak convergence of n_h and p_h and the strong convergence of W_h and ∇_hW_h),

$$-\int_0^T\int_\Omega \overline{n^2}\varphi_t-\overline{n^2}\nabla W\cdot\nabla\varphi\,dxdt\le\int_0^T\int_\Omega\left(\overline{n^2\Delta W}+2\overline{n^2\boldsymbol{G}(p)}\right)\varphi\,dxdt,\tag{4.24}$$

where $\overline{n^2}$ denotes the weak limit of n_h^2 and $\overline{n^2\Delta W}$ and $\overline{n^2\boldsymbol{G}(p)}$ are the weak limits of $n_h^2\Delta_hW_h$ and $n_h^2\boldsymbol{G}(p_h)$ respectively.

Adding (3.9) and (4.24), we have

$$\begin{aligned}
-\int_0^T\int_\Omega\left(\overline{n^2}-n^2\right)\varphi_t-\left(\overline{n^2}-n^2\right)\nabla W\cdot\nabla\varphi\,dxdt&\\
\le\int_0^T\int_\Omega\left(2\overline{n^2\boldsymbol{G}(p)}-2n\overline{n\boldsymbol{G}(p)}+\overline{n^2\Delta W}-n^2\Delta W\right)\varphi\,dxdt.&
\end{aligned}$$

We now choose smooth test functions φ_ϵ approximating $\varphi(t,x)=\mathbf{1}_{[0,\tau]}(t)$, where $\tau\in(0,T]$, in this inequality and then pass to the limit $\epsilon\to 0$ to obtain

$$\int_\Omega \Big(\overline{n^2} - n^2\Big)(\tau)dx - \int_\Omega \Big(\overline{n^2}(0,x) - n^2(0,x)\Big)dx \leq \int_0^\tau \int_\Omega \Big(2\overline{n^2 \boldsymbol{G}(p)} - 2n\overline{n\boldsymbol{G}(p)} + \Delta W\left(\overline{n^2} - n^2\right) + \overline{n^2\Delta W} - \overline{n^2}\Delta W\Big)dx\,dt. \tag{4.25}$$

By convexity of $f(x) = x^2$, we have $\overline{n^2} \geq n^2$, on the other hand, the discrete L^2-entropy inequality, (4.23), implies

$$\int_\Omega |n_h(\tau,x)|^2\,dx \leq \int_\Omega |n_h^0|^2 dx + \int_0^\tau \int_\Omega \left(|n_h|^2 \Delta_h W_h + 2|n_h|^2 \boldsymbol{G}(p_h)\right) dxdt,$$

which gives, passing to the limit $h \to 0$,

$$\int_\Omega \overline{|n|^2}(\tau,x)\,dx \leq \int_\Omega |n_0|^2 dx + \int_0^\tau \int_\Omega \left(\overline{|n|^2\Delta W} + 2\overline{|n|^2\boldsymbol{G}(p)}\right) dxdt.$$

Letting $\tau \to 0$, the second term on the right hand side vanishes (as the integrand is bounded), and we obtain

$$\int_\Omega \overline{|n|^2}(0,x)\,dx \leq \int_\Omega |n_0|^2 dx$$

We deduce that $\overline{|n|^2}(0,\cdot) = |n_0|^2$ almost everywhere and that therefore the second term on the left hand side of (4.25) is zero. We have already estimated the first two terms on the right hand side of (4.25) in (3.13) and (3.14). To bound the other term, we use a discretized version of Lemma 3.7:

LEMMA 4.8. *The weak limits $(n, W, \overline{p})$ of the sequences $\{(n_h, W_h, p_h)\}_{h>0}$ satisfy for any smooth function $S : \mathbb{R} \to \mathbb{R}$,*

$$\int_\Omega \left(\overline{S(n)\Delta W} - \overline{S(n)}\Delta W\right) dx = \frac{1}{\mu}\int_\Omega \left(\overline{p}\;\overline{S(n)} - \overline{pS(n)}\right) dx \tag{4.26}$$

where $\overline{S(n)\Delta W}$, $\overline{S(n)}$, $\overline{pS(n)}$ are the weak limits of $S(n_h)\Delta_h W_h$, $S(n_h)$ and $p_h S(n_h)$ respectively.

Applying this lemma to the last term in (3.12) with $S(n) = n^2$, we can estimate it by

$$\begin{aligned}\int_0^\tau \int_\Omega \left(\overline{n^2\Delta W} - \overline{n^2}\Delta W\right) dx &= \frac{1}{\mu}\int_\Omega \left(\overline{p}\;\overline{n^2} - \overline{pn^2}\right) dxdt \\ &= \frac{1}{\mu}\int_\Omega \left(\overline{n^\gamma}\;\overline{n^2} - \overline{n^{2+\gamma}}\right) dxdt \\ &\leq 0,\end{aligned}$$

using again that by Exercise 3.37 in [**101**], $\overline{n^\gamma}\;\overline{n^2} \leq \overline{n^{2+\gamma}}$. Thus,

$$\int_\Omega \Big(\overline{n^2} - n^2\Big)(\tau)dx \leq \left(2\alpha + \frac{P_M}{\mu}\right)\int_0^\tau \int_\Omega \Big(\overline{n^2} - n^2\Big)\,dx\,dt.$$

Grönwall's inequality thus implies

$$\int_\Omega \left(\overline{n^2} - n^2\right)(\tau)dx \leq 0$$

By convexity of the function $f(x) = x^2$ we also have $n^2 \leq \overline{n^2}$ almost everywhere and hence

$$\overline{n^2} = n^2$$

almost everywhere in $(0,T) \times \Omega$. Therefore we conclude that the functions n_h converge strongly to n almost everywhere, thus also $\overline{p} = n^\gamma$ and so the limit $(n, W, \overline{p})$ is a weak solution of the equations (1.6).

Proof of Lemma 4.8. We multiply the equation for W_h by $S(n_h)$ and integrate it over the spatial domain Ω,

$$\int_\Omega \mu\Delta_h W_h\, S(n_h) - W_h S(n_h)\, dx = -\int_\Omega p_h S(n_h)\, dx.$$

Passing to the limit $h \to 0$ in the last equation, we obtain

$$\int_\Omega \mu\overline{\Delta W S(n)} - W\overline{S(n)}\, dx = -\int_\Omega \overline{pS(n)}\, dx. \tag{4.27}$$

On the other hand, using $[S(n_h) * \psi_\delta](x)$, where ψ_δ is a smooth mollifier converging to a Dirac measure at zero when δ is sent to zero, as a test function in the weak formulation of the limit equation

$$-\mu\Delta W + W = \overline{p},$$

and passing first to the limit $\delta \to 0$ and then $h \to 0$, we obtain

$$\int_\Omega \mu\Delta W\overline{S(n)} - W\overline{S(n)}\, dx = -\int_\Omega \overline{p}\ \overline{S(n)}\, dx$$

Combining the last identity with (4.27), we obtain (4.26). □

5. Numerical examples

To test the scheme in practice, we compute approximations for the following two examples.

5.1. Gaussian initial data. As a first example, we consider the initial data

$$n_0(x) = \frac{1}{2}\exp\left(-10\left(x_1^2 + x_2^2\right)\right), \tag{5.1}$$

on the domain $\Omega = [-2.5, 2.5]^2$ and $h = 1/64$ with pressure law $p = n^3$ and $\boldsymbol{G}(p) = 1 - p$ and $\mu = 1$. Strictly speaking, these are not homogeneous Neumann boundary conditions, but since the gradient of n_0 near the boundary is very small, this works well in practice.

In Figure 1 we show the approximations at times $t = 0, 1, 2, 4$. We observe that the cell density in the middle first reaches the maximum possible and then starts spreading with a relatively narrow transition region between zero density and maximum density.

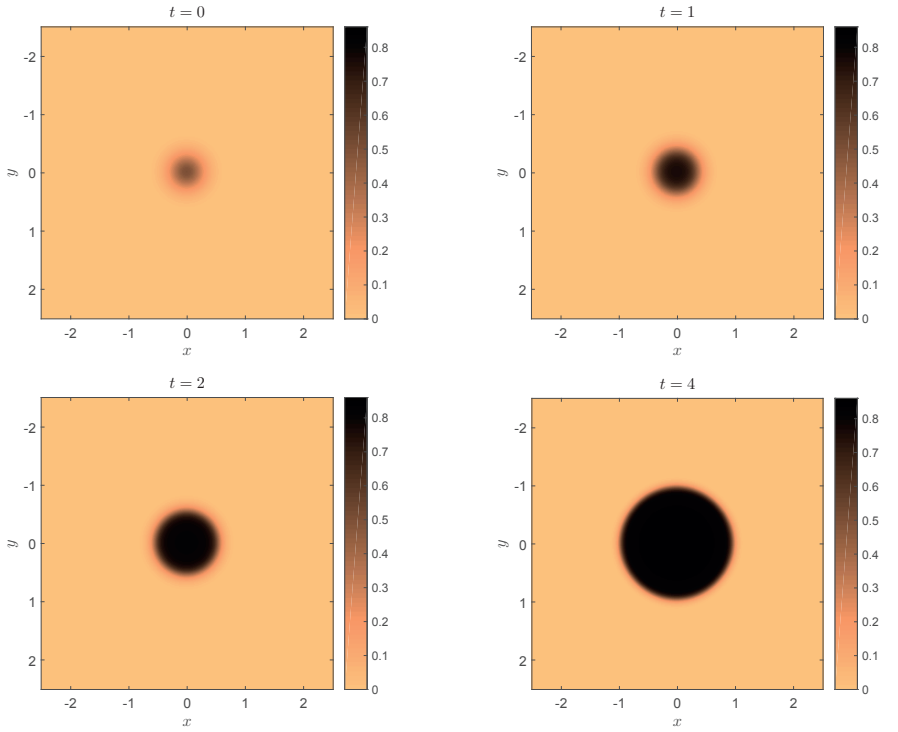

FIGURE 1. The approximations of the cell density n for initial data (5.1) on $\Omega = [-2.5, 2.5]^2$ with mesh width $h = 1/64$.

5.2. Two Gaussians. As a second example, we use the inital data consisting of two Gaussian pulses with centers at $x = (0.7, 0)$ and $x = (-0.6, 0.2)$,

$$n_0(x) = \frac{1}{2}\exp\left(-10\left((x_1 - 0.7)^2 + x_2^2\right)\right) + \frac{1}{2}\exp\left(-20\left((x_1 + 0.6)^2 + (x_2 - 0.2)^2\right)\right) \tag{5.2}$$

on the same domain, $\Omega = [-2.5, 2.5]^2$, with $\mu = 1$, pressure law $p = n^{10}$ and $\boldsymbol{G}(p) = 1 - p$ and mesh width $h = 1/64$. The approximations computed at times $t = 0, 2, 4, 6$ are shown in Figure 2. The interface between the area with maximum cell density and zero cell density seems to be sharper than in the previous example, this appears to be caused by the pressure law with the higher exponent γ. Further tests with higher and lower exponents confirmed that assertion.

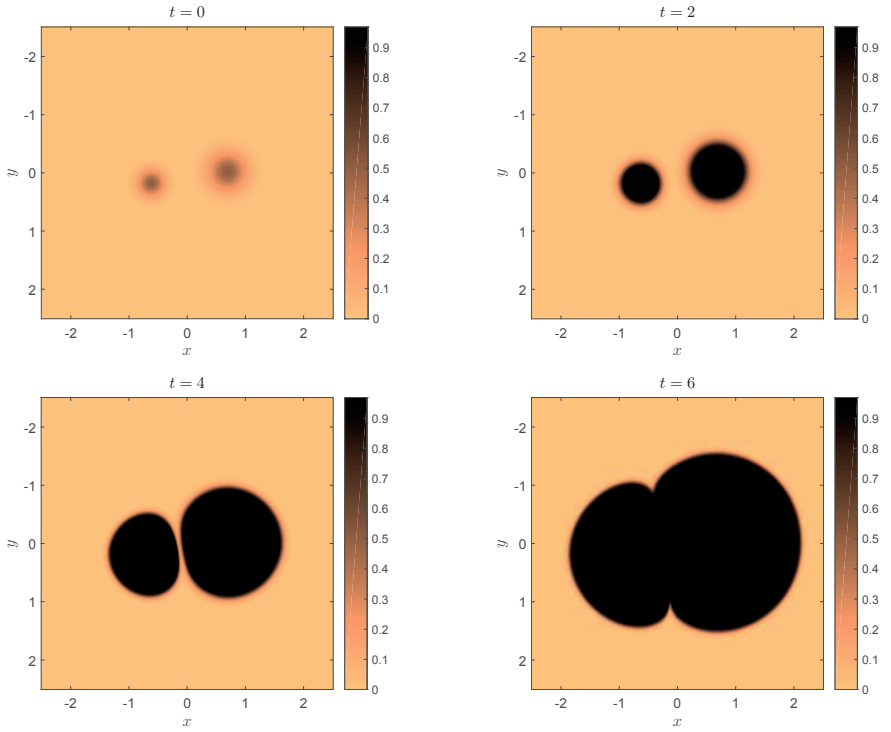

FIGURE 2. The approximations of the cell density n for initial data (5.2) on $\Omega = [-2.5, 2.5]^2$ with mesh width $h = 1/64$.

6. Appendix

6.1. Discretized Aubin-Lions lemma.

LEMMA 6.1. *Let $u_h : \Omega \to \mathbb{R}^k$ be a piecewise constant function defined on a grid on $[0, T) \times \Omega$, Ω a bounded rectangular domain, satisfying*

$$\int_0^T \int_\Omega |u_h|^q + |\nabla_h u_h|^q \, dxdt \leq C \tag{6.1}$$

for some $\infty > q > 1$, uniformly with respect to $h > 0$ and

$$D_t u_h = A_h f_h + g_h + k_h, \tag{6.2}$$

where A_h is a first order linear finite difference operator, and $f_h, g_h, k_h : \Omega \to \mathbb{R}^{d \times k}$ are piecewise constant functions, satisfying uniformly in $h > 0$,

$$\int_0^T \int_\Omega |f_h|^{r_1} + |g_h|^{r_2} + |k_h|\, dxdt \leq C, \tag{6.3}$$

for some $\infty > r_1, r_2 > 1$. *Then* $u_h \to u$ *in* $L^q([0,T) \times \Omega)$.

PROOF. Denote $\widehat{u}_h$ a piecewise linear interpolation of u_h in space piecewise constant in time and similarly, let $\widehat{g}_h$, $\widehat{f}_h$ and $\widehat{k}_h$ piecewise linear interpolations of g_h, f_h and k_h respectively in space and piecewise constant in time such that

$$D_t \widehat{u}_h = A_h \widehat{f}_h + \widehat{g}_h + \widehat{k}_h. \tag{6.4}$$

By Ladyshenskaya's norm equivalences [**88**, p. 230 ff], we have

$$\int_0^T \|\widehat{u}_h\|^q_{W^{1,q}(\Omega)}\, dt \leq C \int_0^T \int_\Omega |u_h|^q + |\nabla_h u_h|^q\, dxdt$$

$$\int_0^T \|\widehat{f}_h\|^{r_1}_{L^{r_1}(\Omega)} + \|\widehat{g}_h\|^{r_2}_{L^{r_2}(\Omega)} + \|\widehat{k}_h\|_{L^1(\Omega)} dt \leq C \int_0^T \int_\Omega |f_h|^{r_1} + |g_h|^{r_2} + |k_h|\, dxdt$$

where the right hand sides are bounded by assumptions (6.1) and (6.3). Since $L^1(\Omega) \subset W^{-1,s}(\Omega)$ for $1 \leq s \leq 1^* = d/(d-1)$, we have that $\widehat{k}_h \in L^1([0,T]; W^{-1,s}(\Omega))$ for $1 \leq s \leq 1^* = d/(d-1)$ and hence thanks to this and (6.4), we obtain

$$\widehat{u}_h \in L^q([0,T); W^{1,q}(\Omega)), \quad D_t \widehat{u}_h \in L^1([0,T); W^{-1,\min\{r_1,1^*\}}(\Omega)),$$

uniformly with respect to the discretization parameter $h > 0$. Thus we can apply the version [**43**, Theorem 1] of the Aubin-Lions lemma to find that up to a subsequence $\widehat{u}_h \to u$ in $L^q([0,T) \times \Omega)$ and the limit $u \in L^q([0,T); W^{1,q}(\Omega))$. By [**88**, Lemma 3.2., p. 226] this implies that also $u_h \to u$ in $L^q([0,T) \times \Omega)$ (and $\nabla_h u_h \rightharpoonup \nabla u$). □

REMARK 6.2 (Derivatives). If the u_h in Lemma 6.1 are of the form $\nabla_h v_h$ for some v_h piecewise constant function, this lemma implies that $\nabla_h v_h \to \nabla v$ in L^q, again applying [**88**, Lemma 3.2., p. 226]

6.2. Technical lemmas. In this section, we prove the following lemma:

LEMMA 6.3. *Let* u_h *solve the difference equation*

$$-\operatorname{div}_h(A_h \nabla_h u_h) + c_h u_h = f_h, \quad x \in \Omega, \tag{6.5}$$

with homogeneous Neumann boundary conditions, where A_h *is a diagonal positive definite* $d \times d$*-matrix with entries* $a_h^{(ii)} \geq \eta > 0$ *and* $c_h \geq \nu > 0$ *uniformly in* $h > 0$, $x \in \Omega$, Ω *is a rectangular domain in* $\mathbb{R}^d$ *and*

$$\|f_h\|_{L^1(\Omega)} \leq M,$$

uniformly in $h > 0$. *We have denoted* $\nabla_h := \nabla_h^-$ *and* $\operatorname{div}_h := \operatorname{div}_h^+$ *(or alternatively* $\nabla_h := \nabla_h^+$ *and* $\operatorname{div}_h := \operatorname{div}_h^-$*). Then*

$$\|u_h\|_{L^q(\Omega)} + \|\nabla_h u_h\|_{L^q(\Omega)} \leq C,$$

where $1 \leq q < d/(d-1)$, *for a constant* $C > 0$ *independent of* $h > 0$.

The proof of this lemma will be a (simplified) finite difference version of the proof of Theorem 2.1 in [**23**]. But before proving the lemma, we need to introduce some notation.

NOTATION 6.4. For any $r \in (1,\infty)$, we denote by $L^{r,\infty}(\Omega)$ the Marcinkiewicz space with norm defined by

$$\|u\|_{L^{r,\infty}(\Omega)} = \sup_{\lambda>0} \lambda |\{x \in \Omega : |u(x) \geq \lambda\}|^{1/r}.$$

The Marcinkiewicz spaces are continuously embedded in $L^q(\Omega)$ for any $1 \leq q < r$, [**53**]:

$$\|u\|_{L^q(\Omega)} \leq C(q,r,|\Omega|)\|u\|_{L^{r,\infty}(\Omega)}, \quad q \in [1,r). \tag{6.6}$$

Moreover, we need the trunctation operator S_k defined as follows:

NOTATION 6.5. Let $k > 0$ be a real number. Then we define the truncation operator $S_k : \mathbb{R} \to \mathbb{R}$ by

$$S_k(s) = \begin{cases} s, & \text{if } |s| \leq k, \\ k\frac{s}{|s|}, & \text{if } |s| \geq k. \end{cases}$$

It will be convenient in the proof to use the following tuple notation for the finite difference approximations:

NOTATION 6.6. We denote $\underline{i} := (i_1,\ldots,i_d)$, $i_\ell = 1,\ldots,N_\ell$, N_ℓ the number of cells in the ℓth spatial direction, a d-dimensional tuple and and $u_{\underline{i}}$ the approximation in cell $\mathcal{C}_{\underline{i}} := ((i_1-1)h, i_1 h] \times \cdots \times ((i_d-1)h, i_d h]$. The piecewise constant function u_h can be written as

$$u_h(x) := \sum_{\underline{i}} u_{\underline{i}} \mathbf{1}_{\mathcal{C}_{\underline{i}}}(x), \quad x \in \Omega.$$

We also need the following auxilary result:

LEMMA 6.7. *Let u_h solve the difference equation* (6.5) *under the assumptions of Lemma* 6.3. *Then*

$$\int_\Omega |\nabla_h S_k(u_h)|^2 + |S_k(u_h)|^2 dx \leq CMk, \quad \forall\, k > 0, \tag{6.7}$$

for some constant $C > 0$ independent of $h > 0$.

PROOF. Given $k > 0$, we multiply equation (6.5) by $S_k(u_h)$ and integrate over the domain Ω. After changing variables in the integrals, we obtain

$$\int_\Omega (A_h \nabla_h u_h) \cdot \nabla_h S_k(u_h) + c_h u_h S_k(u_h)\, dx = \int_\Omega f_h S_k(u_h)\, dx. \tag{6.8}$$

The right hand side can be bounded by Mk using Hölder's inequality. The left hand side, we can rewrite and estimate as follows

$$\begin{aligned}
&\int_\Omega (A_h\nabla_h u_h)\cdot\nabla_h S_k(u_h) + c_h u_h S_k(u_h)\,dx \\
&\quad= \int_\Omega (A_h\nabla_h S_k(u_h))\cdot\nabla_h S_k(u_h) + c_h|S_k(u_h)|^2\,dx \\
&\qquad+ \int_\Omega \left(A_h\left(\nabla_h\left[u_h - S_k(u_h)\right]\right)\right)\cdot\nabla_h S_k(u_h) + c_h\left(u_h - S_k(u_h)\right)S_k(u_h)\,dx \\
&\quad\geq \eta\|\nabla_h S_k(u_h)\|^2_{L^2(\Omega)} + \nu\|S_k(u_h)\|^2_{L^2(\Omega)} \\
&\qquad+ \int_\Omega \left(A_h\left(\nabla_h\left[u_h - S_k(u_h)\right]\right)\right)\cdot\nabla_h S_k(u_h) + c_h\left(u_h - S_k(u_h)\right)S_k(u_h)\,dx.
\end{aligned}$$

$(u_h - S_k(u_h))$ is either zero or has the same sign as $S_k(u_h)$. Therefore $(u_h - S_k(u_h))S_k(u_h) \geq 0$ and

$$\int_\Omega c_h\left(u_h - S_k(u_h)\right)S_k(u_h)\,dx \geq 0.$$

In order to prove that the other term is positive as well, we will show that

$$D^-_\ell S_k(u_{\underline{i}}) D^-_\ell\left(u_{\underline{i}} - S_k(u_{\underline{i}})\right) \geq 0, \quad \forall\,\underline{i},\ \ell = 1,\dots,d.$$

The proof of this fact consists of simple case distinctions and is exactly analoguous for $\ell = 1, 2, (3)$, therefore we will do it only for $\ell = 1$ and omit writing the tuple index $\underline{i}$. Then we have

$$D^-_1(u_i - S_k(u_i))D^-_1 S_k(u_i) = \begin{cases}
(u_i - k)(k - u_{i-1}), & u_i > k,\ |u_{i-1}| \leq k,\\
(u_i + k)(-k - u_{i-1}), & u_i < -k,\ |u_{i-1}| \leq k,\\
0, & |u_i| \leq k,\ |u_{i-1}| \leq k,\\
(-u_{i-1} + k)(u_i - k), & |u_i| \leq k,\ u_{i-1} > k,\\
(-u_{i-1} - k)(u_i + k), & |u_i| \leq k,\ u_{i-1} < -k,\\
0, & u_i > k,\ u_{i-1} > k,\\
0, & u_i < -k,\ u_{i-1} < -k,\\
(u_i - u_{i-1} - 2k)2k, & u_i > k,\ u_{i-1} < -k,\\
-(u_i - u_{i-1} + 2k)2k, & u_i < -k,\ u_{i-1} > k.
\end{cases}$$

The reader is welcome to check that these are all the possible cases and that each of the terms on the right hand side is nonnegative. Thus we have that

$$\int_\Omega (A_h\nabla_h u_h)\cdot\nabla_h S_k(u_h) + c_h u_h S_k(u_h)\,dx \geq \eta\|\nabla_h S_k(u_h)\|^2_{L^2(\Omega)} + \nu\|S_k(u_h)\|^2_{L^2(\Omega)}$$

which implies (6.7) together with the estimate on the right hand side of (6.8) □

PROOF OF LEMMA 6.3. First, we note that by the discrete Gagliardo-Nirenberg-Sobolev inequality [**14**, Thm. 3.4],

$$\int_\Omega |S_k(u_h)|^{2^*} dx \le C^{2^*} \left(\int_\Omega |\nabla_h S_k(u_h)|^2 + |S_k(u_h)|^2 dx \right)^{\frac{2^*}{2}},$$

where $2^* = 2d/(d-2)$ if $d \ge 3$ and any number with $1 \le 2^* < \infty$ if $d = 2$, and where C is a constant depending on $|\Omega|$ but not on $h > 0$. By Lemma 6.7, we can bound the right hand side and obtain therefore

$$\int_\Omega |S_k(u_h)|^{2^*} dx \le C(kM)^{\frac{2^*}{2}}. \tag{6.9}$$

Now we define the set $\mathcal{B}(k)$ by

$$\mathcal{B}(k) = \{\mathcal{C}_{\underline{i}} \subset \Omega : |u_{\underline{i}}| \ge k\}.$$

We have

$$\int_{\mathcal{B}(k)} |S_k(u_h)|^{2^*} dx \ge k^{2^*} |\mathcal{B}(k)|,$$

and therefore, using (6.9),

$$|\mathcal{B}(k)| \le \frac{1}{k^{2^*}} \int_{\mathcal{B}(k)} |S_k(u_h)|^{2^*} dx \le \frac{1}{k^{2^*}} \int_\Omega |S_k(u_h)|^{2^*} dx \le \frac{CM^{\frac{2^*}{2}}}{k^{\frac{2^*}{2}}} \tag{6.10}$$

which implies that $u_h \in L^{r,\infty}(\Omega)$ for $r = 2^*/2$ (which is $d/(d-2)$ if $d \ge 3$) since the choice of $k > 0$ was arbitrary. Now denote

$$\begin{aligned} \partial\mathcal{B}(k) &:= \{\mathcal{C}_{\underline{i}} \subset \Omega : \exists\, \underline{j}, |\underline{i} - \underline{j}| = 1,\ |u_{\underline{j}}| \ge k\} \\ \overline{\mathcal{B}(k)} &:= \mathcal{B}(k) \cup \partial\mathcal{B}(k), \\ \mathcal{B}(k)^c &:= \Omega \backslash \overline{B}(k), \end{aligned}$$

where $|\underline{i} - \underline{j}| = \max_{1 \le \ell \le d} |i_\ell - j_\ell|$. Informally speaking, the cells in $\partial\mathcal{B}(k)$ have a neighbor cell which is contained in $\mathcal{B}(k)$. We have

$$|\partial\mathcal{B}(k)| \le (3^d - 1)|\mathcal{B}(k)| \le \frac{CM^{\frac{2^*}{2}}}{k^{\frac{2^*}{2}}},$$

by (6.10). Now let $\lambda > 0$, $k > 0$ and decompose

$$\begin{aligned} \{x \in \Omega : |\nabla_h u_h(x)| \ge \lambda\} = \{x \in \Omega : |\nabla_h u_h(x)| \ge \lambda \,\text{and}\, x \in \overline{\mathcal{B}(k)}\} \\ \cup \{x \in \Omega : |\nabla_h u_h(x)| \ge \lambda \,\text{and}\, x \in \mathcal{B}(k)^c\}. \end{aligned}$$

Hence

$$|\{x \in \Omega : |\nabla_h u_h(x)| \ge \lambda\}| \le |\overline{\mathcal{B}(k)}| + |\{x \in \Omega : |\nabla_h u_h(x)| \ge \lambda \,\text{and}\, x \in \mathcal{B}(k)^c\}|.$$

On $\mathcal{B}(k)^c$ and the cells bordering the set, we have $|u_h| \le k$ and therefore $u_h = |S_k(u_h)|$. Hence we can estimate the size of the second set in the above inequality,

$$|\{x \in \Omega : |\nabla_h u_h(x)| \ge \lambda \,\text{and}\, x \in \mathcal{B}(k)^c\}|$$

$$
\begin{aligned}
&= |\{x \in \Omega : |\nabla_h S_k(u_h)(x)| \geq \lambda \,\text{and}\, x \in \mathcal{B}(k)^c\}| \\
&\leq |\{x \in \Omega : |\nabla_h S_k(u_h)(x)| \geq \lambda\}| \\
&\leq \frac{1}{\lambda^2} \int_\Omega |\nabla_h S_k(u_h)|^2 dx,
\end{aligned}
$$

where we have used Chebyshev inequality for the last step. Now we can estimate the size of the set $\{x \in \Omega : |\nabla_h u_h(x)| \geq \lambda\}$ using (6.7) once more,

$$
|\{x \in \Omega : |\nabla_h u_h(x)| \geq \lambda\}| \leq \frac{CM^{\frac{2^*}{2}}}{k^{\frac{2^*}{2}}} + \frac{CkM}{\lambda^2}.
$$

Choosing $k = \lambda^{\frac{4}{2^*+2}}$, we obtain

$$
\lambda^{\frac{22^*}{2^*+2}} |\{x \in \Omega : |\nabla_h u_h(x)| \geq \lambda\}| \leq C(d, M, |\Omega|).
$$

If $d \geq 3$, we have $\frac{22^*}{2^*+2} = \frac{d}{d-1}$ and so $u_h, \nabla_h u_h \in L^{r,\infty}(\Omega)$ for $1 \leq r \leq d/(d-1)$. For $d = 2$, since 2^* is an arbitrary finite positive number, we can achieve the same. Using the embedding of the Marcinkiewicz spaces, (6.6), we obtain the claim of the lemma. □

Acknowledgments

The work of K.T. was supported in part by the National Science Foundation under the grant DMS-1211519. The work of F.W. was supported by the Research Council of Norway, project 214495 LIQCRY. F.W. gratefully acknowledges the support by the Center for Scientific Computation and Mathematical Modeling at the University of Maryland where part of this research was performed during her visit in Fall 2014.

PAPER 4

Multilevel Monte Carlo Front Tracking for Random Scalar Conservation Laws

Joint work with Nils Henrik Risebro and Christoph Schwab

Abstract. We consider random scalar hyperbolic conservation laws in spatial dimension $d \geq 1$ with bounded random flux functions which are Lipschitz continuous with respect to the state variable, for which there exists a unique random entropy solution. We present a convergence analysis of a Multilevel Monte Carlo Front Tracking algorithm. It is based on "pathwise" application of the Front Tracking Method for deterministic conservation laws. Due to the first order convergence of front tracking, we obtain an improved complexity estimate in one space dimension.

1. Introduction

Many problems in physics and engineering are modeled by hyperbolic systems of conservation or balance laws. The Cauchy problem for such systems takes the form

$$\begin{gathered} \mathbf{U}_t + \sum_{j=1}^{d} \frac{\partial}{\partial x_j}(\mathbf{F}_j(\mathbf{U})) = 0 \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d,\ t > 0, \\ \mathbf{U}(x,0) = \mathbf{U}_0(x), \quad x \in \mathbb{R}^d. \end{gathered} \tag{1.1}$$

Here, $\mathbf{U} : \mathbb{R}^d \mapsto \mathbb{R}^m$ is the vector of unknowns and $\mathbf{F}_j : \mathbb{R}^m \mapsto \mathbb{R}^m$ is the flux vector for the j-th direction with m being a positive integer.

This type of partial differential equations are ubiquitous, we mention only the shallow water equations of hydrology, the Euler equations for inviscid, compressible flow and the magnetohydrodynamic (MHD) equations of plasma physics, see, e.g. [**37, 58**]. In this paper we focus on the case $m = 1$ in (1.1) which is then called a *scalar conservation law* (SCL).

Solutions of (1.1) develop discontinuities in finite time even when the initial data is smooth. Therefore (1.1) must be interpreted in the weak sense. In order to get uniqueness, (1.1) must be augmented with *entropy conditions*, which at least for scalar conservation laws, makes the initial value problem well-posed. The well-posedness of the Cauchy problem for scalar conservation laws in several space dimensions ($m = 1$, $d \geq 1$) was first established by Kružkov [**85**].

For systems ($m > 1$), some well-posedness results for systems in one space dimension exist [**15, 16**] , but no well-posedness results for systems of conservation laws are available in several space dimensions.

Numerical methods for approximating entropy solutions of systems of conservation laws have undergone extensive development and many efficient methods are available, see [**48, 58, 59, 91**] and the references there. In particular, finite volume methods are frequently employed for approximating (1.1).

This *classical* paradigm for designing efficient numerical schemes assumes that *data for the SCL* (1.1)*, i.e., initial data* $\mathbf{U}_0$ *and flux are known exactly.*

In many situations of practical interest, however, these data are not known exactly due to inherent uncertainty in modelling and measurements of physical parameters such as, for example, the specific heats in the equation of state for compressible gases, resistivity in MHD etc. Often, the initial data are known only up to certain statistical quantities of interest like the mean, variance, higher moments, and in some cases, the law of the stochastic initial data. In such cases, a mathematical formulation of (1.1) is required which allows for *random data*. The problem of random initial data was considered in [**97**], and the existence and uniqueness of a random entropy solution was shown, and a convergence analysis for MLMC FV discretizations was given. Efficient MLMC discretization of balance laws with random source terms was investigated in [**98**].

We mention that the present work as well as [**97, 98**] consider *correlated random inputs* which typically occur in engineering applications; SCLs with random inputs have been considered before, but generally with *white noise*, that is, spatially and temporally uncorrelated random inputs, see [**72, 71, 44, 125, 126**].

In [**97**] a mathematical framework was outlined for deterministic scalar conservation laws with random initial data. This framework was extended to include random flux functions in [**96**]. Here, we generalize [**96**] regarding the existence and uniqueness of random entropy solutions for such problems. We also obtain convergence rate estimates for the approximation of the random entropy solution's expectation by combined front tracking and Monte Carlo sampling.

Specifically, we propose and analyze a multilevel combination of Monte Carlo (MC) sampling the random flux combined with a "pathwise" Front Tracking (FT) solver introduced by Dafermos [**36**] and analyzed, for example, in [**73**], to approximate random entropy solutions of scalar, nonlinear hyperbolic conservation laws.

As the stochastic collocation FVM discretization, and the MLMC FVM algorithms developed in [**98**] also for the numerical solution of nonlinear, hyperbolic *systems* (1.1), the multilevel version of the Monte Carlo Front Tracking method is "non-intrusive" (i.e., it requires only repeated application of existing solvers for input data samples), easy to code and to parallelize, and well-suited for random solutions with low spatial regularity, a situation which is typical in nonlinear hyperbolic conservation laws where discontinuities in realizations of solutions are well known to be generic.

One of our results, Theorem 4.17 and its Corollary 4.18, imply that in space dimension $d = 1$, that the presently proposed MLMC-FT scheme converges in terms of error vs. work with (up to logarithmic terms) at the same rate as one FT solve for the deterministic scalar conservation law; this is stronger than what could be established for MLMC versions of first order finite volume methods in [**99**] and the references there, where first order convergence had to be postulated.

The remainder of this paper is organized as follows: In Section 2, we introduce some preliminary notions from probability theory and functional analysis. The concept of random entropy solutions is introduced and the well-posedness of the scalar hyperbolic conservation law (i.e., (1.1) with $m = 1$) with random initial data is recapitulated in Section 3. The MLMCFT schemes are presented and analyzed in Section 4. Numerical experiments are presented in Section 5.

2. Preliminaries

To set the notation, we recapitulate prerequistes from measure and probability theory which are needed in the subsequent sections. For proofs and further details, we refer for example to [**122**, Chapter 1] and to the references there.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, and let E be a Banach space. A map $G : \Omega \to E$ is called a $\mathbb{P}$-simple function if it is of the form

$$G(\omega) = \sum_{j=1}^{J} g_j \mathbb{1}_{A_j}(\omega), \text{ where } \mathbb{1}_A(\omega) = \begin{cases} 1 & \omega \in A, \\ 0 & \text{otherwise,} \end{cases}$$

and $g_j \in E$ for $j = 1, \ldots, J$, for some finite J and for $A_j \in \mathcal{F}$. A map $f : \Omega \to E$ is strongly $\mathcal{F}$-measurable if there exists a sequence of simple functions f_n converging to f (in the norm of E) $\mathbb{P}$-almost everywhere on Ω.

We call two strongly $\mathbb{P}$-measurable functions $f, g : \Omega \to E$ which agree $\mathbb{P}$-almost everywhere on Ω, $\mathbb{P}$-*versions* of each other. We shall need the following lemma.

LEMMA 2.1. [**122**, Corollary 1.13] *Let E_1 and E_2 be Banach spaces, and $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space. If $f : \Omega \to E_1$ is strongly measurable, and $\phi : E_1 \to E_2$ is continuous, then the composition $\phi \circ f : \Omega \to E_2$ is strongly measurable.*

We define the integral of a simple function $G = \sum g_j \mathbb{1}_{A_j}$ by

$$\int_\Omega G \, d\mathbb{P} = \sum_{j=1}^{N} g_j \mathbb{P}(A_j) \, .$$

If $f : \Omega \to E$ is strongly measurable, we say that f is *Bochner integrable* if there exists a sequence of simple functions $\{f_n\}_{n \geq 0}$ converging to f $\mathbb{P}$-almost everywhere, and

$$\lim_{n \to \infty} \int_\Omega \|f - f_n\|_E \, d\mathbb{P} = 0,$$

([**122**, Def. 1.15]). We then define the Bochner integral of f by

$$\int_\Omega f \, d\mathbb{P} := \lim_{n \to \infty} \int_\Omega f_n \, d\mathbb{P}. \tag{2.1}$$

A strongly measurable function $f : \Omega \to E$ is Bochner integrable if and only if

$$\int_\Omega \|f\|_E \, d\mathbb{P} < \infty$$

(see for example [**122**, Prop. 1.16]) in which case

$$\left\| \int_\Omega f \, d\mathbb{P} \right\|_E \leq \int_\Omega \|f\|_E \, d\mathbb{P} \, . \tag{2.2}$$

For each $1 \leq p < \infty$ we can define the Banach spaces $L^p(\Omega; E)$ to consist of those strongly measurable functions f for which the integrals

$$\int_\Omega \|f\|_E^p \, d\mathbb{P} < \infty \, .$$

These spaces have the natural norm

$$\|f\|_{L^p(\Omega;E)} = \Big(\int_\Omega \|f\|_E^p \, d\mathbb{P} \Big)^{1/p} \, .$$

If $p = \infty$, we define $L^\infty(\Omega; E)$ to be the space of strongly measurable functions $f : \Omega \to E$ for which there exists a number $r \geq 0$ such that $\mathbb{P}(\|f\|_E > r) = 0$. Together with the norm

$$\|f\|_{L^\infty(\Omega;E)} := \inf\{r \geq 0 : \mathbb{P}(\|f\|_E > r) = 0\},$$

this space is a Banach space as well.

If $f : \Omega \to E$ is strongly measurable and $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, we call f an *E-valued random variable*.

3. Hyperbolic conservation laws with random flux

We review classical results on SCLs with deterministic data, and develop a theory of random entropy solutions for SCLs with a class of random flux flunctions, proving in particular the existence and uniqueness of a random entropy solution with finite second moments.

3.1. Deterministic scalar hyperbolic conservation laws. We consider the Cauchy problem for scalar conservation laws (SCL) by setting $m = 1$ in (1.1) and obtaining the SCL in strong form

$$\frac{\partial u}{\partial t} + \sum_{j=1}^{d} \frac{\partial}{\partial x_j} \left(f_j(u) \right) = 0, \quad x = (x_1, \ldots, x_d) \in \mathbb{R}^d, \; t > 0 \, . \tag{3.1}$$

Here the unknown is $u : \mathbb{R}^d \mapsto \mathbb{R}$. Introducing the flux function $f(u)$

$$f(u) = (f_1(u), \ldots, f_d(u)) \in C^1(\mathbb{R}; \mathbb{R}^d) \, , \quad \operatorname{div} f(u) = \sum_{j=1}^{d} \frac{\partial}{\partial x_j} f_j(u),$$

we may rewrite (3.1) succinctly as

$$\frac{\partial u}{\partial t} + \operatorname{div} \left(f(u) \right) = 0 \;\; \text{for } (x, t) \in \mathbb{R}^d \times \mathbb{R}_+. \tag{3.2}$$

We supply the SCL (3.2) with initial condition

$$u(x, 0) = u_0(x), \;\; x \in \mathbb{R}^d \, . \tag{3.3}$$

3.2. Entropy solutions. Solutions to (3.1) are in general not smooth since they can develop discontinuities in finite time. Therefore we look for weak solutions to the equations. In particular, we are interested in distributional solutions in the class of *entropy solutions* which satisfy in addition the entropy condition

$$\eta(u)_t + \operatorname{div} Q(u) \leq 0, \quad \text{in } \mathcal{D}(\mathbb{R}^d \times \mathbb{R}^+),$$

for all entropy pairs (η, Q), where η, the *entropy*, is a convex C^2-function and $Q(u) = (Q_1(u), \dots, Q_d(u))$, the *entropy flux*, satisfies $Q_j' = \eta' f_j'$. In this class, uniqueness can be proved [**85**]. We will in the following restrict to initial data in $L^\infty(\mathbb{R}^d) \cap BV(\mathbb{R}^d)$, but results can be proved for more general initial conditions [**103**]. By a *function of bounded variation*, or BV-function, we mean a function $f \in L^1(\mathbb{R}^d)$ with

$$TV(f) := \sup\left\{ \int_{\mathbb{R}^d} f \operatorname{div} \varphi \, dx \;\middle|\; \varphi \in C_0^1(\mathbb{R}^d; \mathbb{R}^d),\, |\varphi| \leq 1 \right\} < \infty,$$

where $|\varphi|$ denotes absolute value of point-values for φ, see [**47**, Section 5.1]. We call $TV(f)$ the *total variation* of f. We define the Banach space of functions with bounded variation as the completion of $C_0^\infty(\mathbb{R}^d)$ with respect to the norm

$$\|f\|_{BV(\mathbb{R}^d)} := \|f\|_{L^1(\mathbb{R}^d)} + TV(f).$$

More details and properties of BV-functions can be found in, for example [**47**, Chapter 5], [**73**, Appendix A] or [**57**, Chapter 1]. Next we introduce the (nonlinear) data-to-solution operator

$$S_t : (u_0, f) \longmapsto u(\cdot, t) =: S_t(u_0, f) \quad t > 0\,.$$

In particular, we shall need the following continuity (with respect to initial data and flux function) result for deterministic scalar conservation laws:

THEOREM 3.1. [**73**, Thm. 2.14, Thm. 4.3] *Assume $u_0, v_0 \in (BV \cap L^\infty)(\mathbb{R}^d)$, and f, $g \in \operatorname{Lip}(\mathbb{R}; \mathbb{R}^d)$. Then there exist unique entropy solutions u and v to* (3.1) *with initial data u_0 and v_0 respectively and flux functions f and g, which satisfy the a-priori continuity estimates: For all $t \geq 0$ we have*

$$\|u(\cdot,t) - v(\cdot,t)\|_{L^1(\mathbb{R}^d)} \leq \|u_0 - v_0\|_{L^1(\mathbb{R}^d)} + t \min\{\mathrm{TV}(u_0), \mathrm{TV}(v_0)\} \|f - g\|_{W^{1,\infty}(\mathbb{R};\mathbb{R}^d)}, \tag{3.4}$$

and

$$\|u(\cdot,t) - u(\cdot,s)\|_{L^1(\mathbb{R}^d)} \leq (t-s)\, \mathrm{TV}(u_0)\, \|f\|_{W^{1,\infty}(\mathbb{R},\mathbb{R}^d)}\,, \tag{3.5}$$

for all $0 \leq s \leq t$. In particular, this implies that the soluiton operator S_t is a uniformly continuous mapping from $BV(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d) \times W^{1,\infty}_{\mathrm{loc}}(\mathbb{R})$ into $C([0,T]; L^1(\mathbb{R}^d))$. Moreover, it follows that

$$\|u(\cdot,t) - v(\cdot,t)\|_{L^1(\mathbb{R}^d)} \leq \|u_0 - v_0\|_{L^1(\mathbb{R}^d)}\,. \tag{3.6}$$

if $f \equiv g$, and

$$\mathrm{TV}(u(\cdot,t)) \leq \mathrm{TV}(u_0), \tag{3.7}$$

$$\|u(\cdot,t)\|_{L^\infty(\mathbb{R}^d)} \leq \|u_0\|_{L^\infty(\mathbb{R}^d)}, \tag{3.8}$$

$$\|u(\cdot,t)\|_{L^1(\mathbb{R}^d)} \leq \|u_0\|_{L^1(\mathbb{R}^d)}. \tag{3.9}$$

For a proof, we refer to for example [**73**, Theorem 2.14 and Theorem 4.3], or other standard references such as [**58, 59, 48, 91, 103**].

3.3. Random flux and initial data. Existence and uniqueness in the case of random initial data u_0 and continuously differentiable random flux f was proved in [**97, 96**]. Here, we are interested in initial data u_0 and flux functions f_j in (3.1) which are random elements with values in $BV(\mathbb{R}^d)\cap L^\infty(\mathbb{R}^d)$ and $W^{1,\infty}(\mathbb{R};\mathbb{R})$ respectively. To define these, we denote by $(\Omega,\mathcal{F},\mathbb{P})$ a probability space. We consider *spatially homogeneous random flux functions* f, i.e., strongly measurable maps $f:\Omega\to \mathrm{Lip}(\mathbb{R};\mathbb{R}^d)$, and random initial data u_0 being strongly measurable maps from Ω to the intersection of the Banach spaces $BV(\mathbb{R}^d)$ and $L^\infty(\mathbb{R}^d)$.

Definition 3.2. Random data for the SCL (3.1) is a random variable taking values in

$$E_1 = \big(BV(\mathbb{R}^d)\cap L^\infty(\mathbb{R}^d)\big)\times W^{1,\infty}(\mathbb{R};\mathbb{R}^d).$$

The set E_1 is a Banach space which we equip with the norm

$$\|(u,f)\|_{E_1} = \|u\|_{L^1(\mathbb{R}^d)} + \mathrm{TV}(u) + \|u\|_{L^\infty(\mathbb{R}^d)} + \|f\|_{W^{1,\infty}(\mathbb{R};\mathbb{R}^d)}\ . \tag{3.10}$$

In particular, random data (u_0,f) for the SCL (3.1) - (3.3) is a strongly measurable map

$$(u_0,f):(\Omega,\mathcal{F})\longmapsto (E_1,\mathcal{B}(E_1))\ . \tag{3.11}$$

For the ensuing convergence analysis, we shall also require that

$$\|u_0\|_{L^\infty(\Omega;(L^\infty\cap BV)(\mathbb{R}^d))}\leq \overline{M}<\infty,\ \text{and}\ \ \|f\|_{L^\infty(\Omega;W^{1,\infty}([-\overline{M},\overline{M}];\mathbb{R}^d))}\leq \overline{M}<\infty\,. \tag{3.12}$$

We shall refer to a random flux f which satisfies (3.12) as *bounded random flux*. By (2.2), for random data with (3.12) the map

$$\Omega\ni\omega\mapsto \Big(\|u_0(\omega;\cdot)\|_{L^1(\mathbb{R}^d)}, \mathrm{TV}(u_0(\omega;\cdot)), \|u_0(\omega;\cdot)\|_{L^\infty(\mathbb{R}^d)}, \|f\|_{W^{1,\infty}(\mathbb{R};\mathbb{R}^d)}\Big) \tag{3.13}$$

is in $L^k(\Omega;\mathbb{R}^4)$ for every $1\leq k<\infty$.

3.4. Random entropy solution. Based on Theorem 3.1, we formulate (3.1) - (3.3) for random data (u_0,f) in the sense of Definition 3.2. We are interested in solutions of the *random scalar conservation law* (RSCL)

$$\begin{cases}\partial_t u(\omega;x,t)+\mathrm{div}_x(f(\omega;u(\omega;x,t)))=0, & t>0,\\ u(\omega;x,0)=u_0(\omega;x), & x\in\mathbb{R}^d.\end{cases} \tag{3.14}$$

Definition 3.3. A random variable $u:\Omega\ni\omega\to u(\omega;x,t)$, i.e., a strongly measurable mapping from $(\Omega,\mathcal{F})$ to $C([0,T];L^1(\mathbb{R}^d))$, is a *random entropy solution* of the SCL (3.14) with random data as in (3.11) - (3.13) for some $k\geq 2$, if u satisfies the following:

(i.) Weak solution: For $\mathbb{P}$-a.e $\omega \in \Omega$, $u(\omega;\cdot,\cdot)$ satisfies

$$\int_0^\infty \int_{\mathbb{R}^d} \Big(u(\omega;x,t)\varphi_t(x,t) + \sum_{j=1}^d f_j(\omega;u(\omega;x,t))\frac{\partial}{\partial x_j}\varphi(x,t)\Big)\,dxdt + \int_{\mathbb{R}^d} u_0(x,\omega)\varphi(x,0)\,dx = 0, \tag{3.15}$$

for all test functions $\varphi \in C_0^\infty(\mathbb{R}^d \times \mathbb{R})$.

(ii.) Entropy condition: For any pair consisting of a (deterministic) entropy η and a (stochastic) entropy flux $Q(\omega;\cdot)$ i.e., η, Q_j with $j = 1, 2, \ldots, d$ are functions such that η is convex and such that $Q_j'(\omega;\cdot) = \eta' f_j'(\omega;\cdot)$ for all j, and for $\mathbb{P}$-a.e $\omega \in \Omega$, u satisfies the following integral identity,

$$\int_0^\infty \int_{\mathbb{R}^d} \Big(\eta(u(\omega;x,t))\varphi_t(x,t) + \sum_{j=1}^d Q_j(\omega;u(\omega;x,t))\frac{\partial}{\partial x_j}\varphi(x,t)\Big)\,dxdt + \int_{\mathbb{R}^d} \eta(u_0(\omega;x))\varphi(x,0)\,dx \geq 0, \tag{3.16}$$

for all non-negative test functions $\varphi \in C_0^\infty(\mathbb{R}^d \times \mathbb{R})$.

THEOREM 3.4. *Consider the SCL* (3.1) - (3.3) *with random data* (u_0, f) *in the sense of Definition* 3.2 *such that* (3.12) *holds. Then there exists a random entropy solution* u *in* $C([0,T];L^1(\mathbb{R}^d))$*, which for each* $0 \leq t \leq T$ *is described by the map*

$$\Omega \ni \omega \mapsto u(\omega;\cdot,t) = S_t(u_0(\omega,\cdot), f(\omega;\cdot))\ .$$

For $\mathbb{P}$*-almost every* $\omega \in \Omega$ *we have the bound*

$$\|u(\omega;\cdot,t)\|_{(L^\infty \cap BV)(\mathbb{R}^d)} \leq \|u_0(\omega;\cdot)\|_{(L^\infty \cap BV)(\mathbb{R}^d)}\,, \tag{3.17}$$

and for all $k \geq 1$, $(u_0, f) \in L^k(\Omega; E_1)$ *implies that*

$$\|u\|_{L^k(\Omega;C([0,T];L^1(\mathbb{R}^d)))} \leq \|(u_0,f)\|_{L^k(\Omega;E_1)}\,. \tag{3.18}$$

PROOF. Let $E_2 = C([0,T], L^1(\mathbb{R}^d))$. By (3.12), for almost all ω, the data $u_0(\omega;\cdot)$ and $f(\omega;\cdot)$ are such that there exists a unique entropy solution $u(\omega;\cdot) \in E_2$ to (3.14). Furthermore, from (3.7) – (3.9) it follows that for such ω,

$$\|u(\omega;\cdot,t)\|_{(L^\infty \cap BV)(\mathbb{R}^d)} \leq \|u_0(\omega;\cdot)\|_{(L^\infty \cap BV)(\mathbb{R}^d)}\ ,$$

which implies (3.17). We have to show that $\omega \mapsto u(\omega;\cdot)$ is a random variable, that is, it is strongly measurable. This will follow from Lemma 2.1 if the mapping $E_1 \ni (u_0, f) \mapsto u \in E_2$ is continuous. This on the other hand, follows from (3.4) and (3.5) in Theorem 3.1.

To prove (3.18), we compute

$$\begin{aligned}\|u\|^k_{L^k(\Omega;C([0,T];L^1(\mathbb{R}^d)))} &= \int_\Omega \sup_{t\le T} \|u(\omega;\cdot,t)\|^k_{L^1(\mathbb{R}^d)}\, d\mathbb{P} \\ &\le \int_\Omega \|u_0(\omega;\cdot)\|^k_{L^1(\mathbb{R}^d)}\, d\mathbb{P} \\ &\le \|(u_0,f)\|^k_{L^k(\Omega;E_1)}\,.\end{aligned}$$

□

REMARK 3.5. The random entropy solution $u : \Omega \to C([0,T];L^1(\mathbb{R}^d))$ is unique in the sense that if a random variable $(\widetilde{u}_0, \widetilde{f})$ is a $\mathbb{P}$-version of (u_0, f), then the solution $\widetilde{u}(\cdot;\cdot,t) := S_t(\widetilde{u}_0, \widetilde{f})$ corresponding to it is a $\mathbb{P}$-version of $u(\cdot;\cdot,t) := S_t(u_0, f)$, that is, they agree everywhere on Ω except on a set with $\mathbb{P}$-measure zero. To see this, we note that by the continuity of the operator S_t, (3.4), we have for any $t \in (0,T]$,

$$\begin{aligned}&\|u(\cdot;\cdot,t) - \widetilde{u}(\cdot;\cdot,t)\|_{L^\infty(\Omega;L^1(\mathbb{R}^d))} \\ &\quad\le \|\widetilde{u}_0 - u_0\|_{L^\infty(\Omega;L^1(\mathbb{R}^d))} \\ &\qquad + t\min\{\|TV(u_0)\|_{L^\infty(\Omega)}, \|TV(\widetilde{u}_0)\|_{L^\infty(\Omega)}\}\|f-\widetilde{f}\|_{L^\infty(\Omega;W^{1,\infty}([-\overline{M},\overline{M}];\mathbb{R}^d))} \\ &\quad= 0,\end{aligned}$$

and therefore it follows also that u is unique in $L^\infty(\Omega;L^1(\mathbb{R}^d);d\mathbb{P})$.

4. Multilevel Monte Carlo front tracking

In this section, we present a Multilevel Monte Carlo (MLMC) version of the front tracking approach to the numerical solution of hyperbolic conservation laws with random flux (3.15), (3.16) as developed in [**73**].

4.1. The Monte Carlo Method. We interpret the Monte Carlo method as "discretization" of the SCL random data $f(\omega;u)$, $u_0(\omega;x)$ as in (3.11) – (3.13) with respect to ω. We assume in particular the existence of k-th moments of u_0 for some $k \in \mathbb{N}$. We shall be interested in the statistical estimation of the first and higher moments of u, i.e, $\mathcal{M}^k(u) \in (L^1(\mathbb{R}^d))^{(k)}$. For $k=1$, $\mathcal{M}^1(u) = \mathbb{E}[u]$. The *MC approximation of* $\mathbb{E}[u]$ is defined as follows: given M independent, identically distributed samples $(\widehat{u}^i_0, \widehat{f}^i)$, $i = 1,\dots,M$, of random data, the MC estimate of $\mathbb{E}[u(\cdot;\cdot,t)]$ at time t is

$$E_M[u(\cdot,t)] := \frac{1}{M}\sum_{i=1}^M \widehat{u}^i(\cdot,t) \tag{4.1}$$

where $\widehat{u}^i(\cdot,t)$ denotes the M unique entropy solutions of the M Cauchy Problems (3.1) – (3.3) with initial data $\widehat{u}^i_0$ and flux samples $\widehat{f}^i(\cdot)$. We observe that by

$$\widehat{u}^i(\cdot,t) = S_t(\widehat{u}^i_0, \widehat{f}^i)$$

we have for every M and for every $0 < t < \infty$, by (3.9),

$$\begin{aligned}\|E_M[u(\omega;\cdot,t)]\|_{L^1(\mathbb{R}^d)} &= \Big\|\frac{1}{M}\sum_{i=1}^{M} S_t((\widehat{u}_0^i,\widehat{f}^i)(\omega))\Big\|_{L^1(\mathbb{R}^d)}\\ &\le \frac{1}{M}\sum_{i=1}^{M}\Big\|S_t((\widehat{u}_0^i,\widehat{f}^i)(\omega))\Big\|_{L^1(\mathbb{R}^d)}\\ &\le \frac{1}{M}\sum_{i=1}^{M}\big\|\widehat{u}_0^i(\omega;\cdot)\big\|_{L^1(\mathbb{R}^d)}.\end{aligned}$$

Using the i.i.d. property of the samples $\{\widehat{u}_0^i,\widehat{f}^i\}_{i=1}^M$, Theorem 3.4 and and the linearity of the expectation $\mathbb{E}[\cdot]$, we obtain the bound

$$\mathbb{E}\left[\|E_M[u(\cdot;\cdot,t)]\|_{L^1(\mathbb{R}^d)}\right] \le \mathbb{E}\left[\|u_0\|_{L^1(\mathbb{R}^d)}\right] = \|u_0\|_{L^1(\Omega;L^1(\mathbb{R}^d))} < \infty.$$

As $M\to\infty$, the MC estimates (4.1) converge and the convergence result from [**97**] holds as well.

THEOREM 4.1. *Assume that in the SCL* (3.1) – (3.3) *the random data* (u_0, f) *satisfies* (3.12).

Then for every $t > 0$ *the MC estimates* $E_M[u(\cdot,t)]$ *in* (4.1) *converge in* $L^2(\Omega;L^1(\mathbb{R}^d))$ *as* $M\to\infty$, *to* $\mathcal{M}^1(u(\cdot,t)) = \mathbb{E}[u(\cdot,t)]$ *and, for any* $M\in\mathbb{N}$, $0<t<\infty$, *we have the error bound*

$$\|\mathbb{E}[u(\cdot,t)] - E_M[u(\cdot,t)]\|_{L^2(\Omega;L^1(\mathbb{R}^d))} \le 2M^{-1/2}\,\|u_0\|_{L^2(\Omega;L^1(\mathbb{R}^d))}\ .$$

4.2. Front tracking. As an exact solution to (3.1) – (3.3) is in general not available, an approximate solution has to be computed numerically. Here, we investigate using a front tracking method described in [**36, 73, 69, 68**]. Since the method and the associated convergence analysis differ for the dimensions $d = 1$ and $d > 1$, we treat the two cases separately.

4.2.1. *Front tracking in the one dimensional case.* We start by briefly describing the front tracking algorithm for the deterministic conservation law (3.1) – (3.3) with initial condition u_0 given in $BV(\mathbb{R})\cap L^\infty(\mathbb{R})$. Let $\overline{M} := \|u_0\|_{L^\infty(\mathbb{R})}$ and let $\delta > 0$ be a small number. Moreover, set $u_i = \delta i$, for $-\overline{M}\le i\delta\le\overline{M}$, and discretize the spatial domain by a grid $\{x_j = j\delta,\, j\in\mathbb{Z}\}$. Then, u_0 is approximated by a piecewise constant function u_0^δ taking in each cell $[j\delta,(j+1)\delta)$ one of the values in $V_\delta := \{u_i \,|\, i\in\mathbb{Z},\ |u_i|\le\overline{M}\}$. The flux function f is approximated by the piecewise linear interpolation f^δ,

$$\begin{aligned} f^\delta(u) &= f(u_j) + \frac{f(u_{j+1})-f(u_j)}{u_{j+1}-u_j}(u-u_j),\\ u&\in[u_j,u_{j+1}),\quad j\in\mathbb{Z},\ |j|\le\overline{M}\delta^{-1}\,.\end{aligned} \tag{4.2}$$

Then we solve the initial value problem

$$u_t^\delta + f^\delta(u^\delta)_x = 0,\quad (x,t)\in\mathbb{R}\times(0,T), \tag{4.3a}$$

$$u^\delta(x,0) = u_0^\delta(x), \quad x \in \mathbb{R}, \tag{4.3b}$$

exactly. This means that in each step, we solve the Riemann problems between the states of the piecewise constant function u^δ, then track the discontinuities, called *fronts*, until they interact, solve the emerging Riemann problem and so on. Note that the solution of each Riemann problem is again a piecewise constant function taking values in V_δ because f^δ is piecewise linear with breakpoints $u_i \in V_\delta$. Thus, the (unique) entropy solution $u^\delta(\cdot,t)$ is a piecewise constant function for all $t > 0$. It was shown in [**73**, Lemma 2.6] that the number of interactions $T(\delta,t)$ between fronts for $t \in (0,\infty)$ is bounded by

$$T(\delta,t) \le \frac{1}{\delta}(|V_\delta|+1)\ \mathrm{TV}(u^\delta) \le \frac{1}{\delta}(2\lceil \overline{M}/\delta \rceil + 1)\ \mathrm{TV}(u^\delta) \tag{4.4}$$

where we denoted $|V_\delta|$ the cardinality of the set V_δ which is bounded for all $t > 0$ by $2\lceil \overline{M}/\delta \rceil$ due to (4.5). Hence the process terminates. Moreover, the solution u^δ of (4.3) satisfies the Kružkov entropy condition and we have the theorem:

THEOREM 4.2 ([**73**]). *For initial data $u_0 \in BV(\mathbb{R}) \cap L^\infty(\mathbb{R})$ and flux function $f(u) \in W^{2,\infty}_{\mathrm{loc}}(\mathbb{R})$ we have*

(i) The solutions u^δ to the differential equation (4.3) *are uniformly bounded in δ for all $t \in (0,\infty)$:*

$$\|u^\delta(\cdot,t)\|_{L^\infty(\mathbb{R})} \le \|u_0\|_{L^\infty(\mathbb{R})}, \quad t \in (0,\infty), \tag{4.5}$$

(ii) The total variation of u^δ is bounded by the total variation of the initial data for all times $t \in (0,\infty)$,

$$\mathrm{TV}(u^\delta(\cdot,t)) \le \mathrm{TV}(u_0), \quad t \in (0,\infty),$$

(iii) As the discretization parameter δ tends to zero, the sequence $\{u^\delta\}_{\delta>0}$ converges in $C([0,T];L^1(\mathbb{R}))$ to the unique entropy solution u of (3.1) – (3.3). *Specifically,*

$$\|u(\cdot,t) - u^\delta(\cdot,t)\|_{L^1(\mathbb{R})} \le \|u_0 - u_0^\delta\|_{L^1(\mathbb{R})} + t\|f - f^\delta\|_{\mathrm{Lip}(\mathbb{R})}\,\mathrm{TV}(u_0) \tag{4.6}$$

COROLLARY 4.3. *Under the assumptions of Theorem* 4.2*, we have the following estimate with respect to the discretization parameter δ:*

$$\|u(\cdot,t) - u^\delta(\cdot,t)\|_{L^1(\mathbb{R})} \le \delta\ \mathrm{TV}(u_0)\,(c + |f|_{W^{2,\infty}(\mathbb{R})})\ . \tag{4.7}$$

PROOF. We note that the regularity $f \in W^{2,\infty}$ implies (e.g. [**73**, Ex. 2.11])

$$\|f - f^\delta\|_{\mathrm{Lip}(\mathbb{R})} \le \delta\|f''\|_{L^\infty(\mathbb{R})} = \delta|f|_{W^{2,\infty}(\mathbb{R})}, \tag{4.8}$$

and that the cell-average approximation u_0^δ of $u_0 \in BV(\mathbb{R}^d)$ satisfies (see [**73**])

$$\|u_0 - u_0^\delta\|_{L^1(\mathbb{R})} \le \delta\, c\ \mathrm{TV}(u_0),$$

where $c > 0$ is independent of δ. For a proof of the latter inequality, consider for example equation (4.30) in [**73**]. Then (4.7) follows using (4.8) and (4.6). □

In order to obtain convergence rate bounds in the Multilevel Monte Carlo front tracking (MCMLFT) algorithm, which we are going to introduce in the next section, it will be useful to have convergence rates of the front tracking algorithm with respect to the amount of computational *work* of the algorithm when the discretization is refined.

DEFINITION 4.4. By the (*computational*) *work* or *cost* of an algorithm, we mean the number of floating point operations performed during the execution of the algorithm. We assume that this is proportional to the run time of the algorithm.

LEMMA 4.5 (Work estimate). *Under the assumptions of Theorem* 4.2*, the front tracking approximation* u^δ *satisfies the following estimate with respect to the total cost* W_δ^{FT} *of the front tracking algorithm,*

$$\|u(\cdot,t) - u^\delta(\cdot,t)\|_{L^1(\mathbb{R})} \leq C\ \mathrm{TV}(u_0) \times \left(1 + \|f\|_{W^{2,\infty}(\mathbb{R})}\right)\left((\|u_0\|_{L^\infty} + 1)\left(\mathrm{TV}(u_0) + \|u_0\|_{L^\infty}\right)\right)^{1/2} \left(W_\delta^{FT}\right)^{-1/2}. \tag{4.9}$$

PROOF. Theorem 4.2 implies in particular that we have for the total number of interactions (4.4), (due to (3.12), in the case of random initial data holds $\mathrm{TV}(u_0) \leq \overline{M}$ $\mathbb{P}$-as.)

$$T(\delta,t) \leq \frac{1}{\delta}(2\lceil \overline{M}/\delta \rceil + 1)\ \mathrm{TV}(u_0^\delta) \leq \frac{C}{\delta^2}(\|u_0\|_{L^\infty(\mathbb{R})} + 1)\ \mathrm{TV}(u_0), \tag{4.10}$$

and that the number of different Riemann problems that might be solved during the execution of the algorithm is bounded by $4\lceil \overline{M}/\delta \rceil^2$. We use Algorithm 1, which is a modification of Graham's scan [**60**] used to compute the convex hull of a set of points in the plane, to calculate all the solutions of the Riemann problems with left state $u_i = i\delta$, right state $u_j = j\delta$, $L \leq i < j \leq R$, where L, R are chosen such that $u_L = \min V_\delta$, $u_R = \max V_\delta$ (a similar algorithm can be used to compute the solutions to the Riemann problems with left state $u_i = i\delta$, right state $u_j = j\delta$, $R \leq j < i \leq L$). It can easily be verified (see [**60**]) that the cost of the execution of Algorithm 1 is bounded by $C\,\overline{M}^2\,\delta^{-2}$, where C is a constant independent of $\overline{M}$ and δ, for the input $\delta > 0$, $L = -\lceil \overline{M}/\delta \rceil$, $R = \lceil \overline{M}/\delta \rceil$.

So, if the solutions to all possible Riemann problems are computed and stored in advance, the work W_δ^{FT} to compute the front tracking approximation $u^\delta(\cdot,t)$ is bounded by $C(\|u_0\|_{L^\infty} + 1)(\mathrm{TV}(u_0) + \|u_0\|_{L^\infty})\,\delta^{-2}$, for a constant $C > 0$, uniformly in $t \in (0,\infty)$. We thus obtain (4.9)

□

REMARK 4.6. Note that the work W_δ^{FT} to compute the front tracking approximation is of the same order as the work we would need to compute an approximation of the solution by a finite volume scheme on a grid with cells of diameter $\mathcal{O}(\delta)$. But due to the better convergence rate with respect to the discretization parameter δ, which is of order 1 whereas it is proved to be of order 1/2 for the finite volume approximation, we obtain the improved convergence rate (4.9) with respect to work.

Algorithm 1 Compute Riemann problems with $u_L \le u_i < u_j \le u_R$

Input: $\delta > 0$, $L < R \in \mathbb{Z}$, (u_L smallest value of u, u_R largest value of u), $\underline{f} = [f_L, \dots, f_R]$, ($f_i = f(u_i)$, $L \le i \le R$)

Output: $U_{i,j} = [u_i, \dots, u_j]$ (states present in solution of RP with left state u_i and right state u_j), $s_{i,j} = [s^1_{i,j}, \dots, s^{k_{ij}}_{i,j}]$ (vector of shock speeds (in increasing order) present in RP with left state u_i and right state u_j, $k_{ij} \in \mathbb{N}$), $L \le i < j \le R$

```
for i = L to R do
    û ← [i, i + 1]
    ŝ ← (f_{i+1} − f_i)/δ
    s_{i,i+1} ← ŝ
    U_{i,i+1} ← δ · û
    k ← i + 2
    while k ≤ R do
        sl ← (f_k − f_{û(end)})/(δ(k − û(end)))
        if ŝ = [ ] or sl > ŝ(end) then
            ŝ ← [ŝ, sl]
            û ← [û, k]
            s_{i,k} ← ŝ
            U_{i,k} ← δ · û
            k ← k + 1
        else
            ŝ ← ŝ(1 : end − 1)
            û ← û(1 : end − 1)
        end if
    end while
end for
```

REMARK 4.7 (Work estimates for convex flux functions). If the flux function f is convex, the work estimate can be improved. This is because in this case, the number of interactions $T(\delta, t)$ can be bounded by the sum of the sizes of the jumps in the initial data. That is, given u_0 there holds, for every $t > 0$ and $\delta > 0$,

$$T(\delta, t) \le \frac{1}{\delta} \operatorname{TV}(u_0)$$

(see [**73**, Lemma 2.6]), since for a convex flux function, the number of fronts is strictly decreasing at each interaction. Moreover, the solution of each Riemann problem is either a shock wave or a rarefaction wave depending on whether $u_L > u_R$ or $u_L < u_R$, and we do not need to compute the convex envelope of the flux function.

So, the solution of one Riemann problem can be computed with a cost proportional to δ. Thus the total work W^{FT}_δ to compute the front tracking approximation reduces to

$$W^{FT}_\delta \le C \ \operatorname{TV}(u_0)\, \delta^{-1}$$

and we obtain the improved convergence rate of the FT method with respect to work,

$$\|u(\cdot,t)-u^\delta(\cdot,t)\|_{L^1(\mathbb{R})} \le C\,\mathrm{TV}(u_0)^2(1+\|f\|_{W^{2,\infty}(\mathbb{R})})\,(W_\delta^{FT})^{-1}. \tag{4.11}$$

Clearly, the same rate holds also for concave fluxes.

4.2.2. *Front tracking for $d \ge 2$ and dimensional splitting.* Front tracking in several space dimensions is based on the method of fractional steps (or dimensional splitting) introduced by Bagrinovskiĭ and Godunov [**6**] and later on extended by various authors, see e.g. [**70**] and the references therein. Here, we will use the dimensional splitting method in combination with the front tracking algorithm for one space dimension as described in the previous Section 4.2.1. To describe the method, we introduce some notation. We discretize the spatial domain by a Cartesian grid $\{j\Delta x_i, j\in\mathbb{Z}\}$, $i=1,\dots,d$ in each direction and denote by $I_{j_1,\dots,j_d}$ the grid cell

$$I_{j_1,\dots,j_d} = \big\{(x_1,\dots,x_d) \;\big|\; j_i\Delta x_i \le x_1 < (j_i+1)\Delta x_i \;\text{ for } i=1,\dots,d\big\}\ .$$

Moreover, we denote the projection operator $\pi_\delta := P_\delta\circ\overline{P}_{\Delta x}$ for a function $u\in L^1(\mathbb{R}^d)$ to be the composition of the projection $\overline{P}_{\Delta x}$ of the function on the cell averages,

$$\overline{P}_{\Delta x}u(x) = \frac{1}{\Delta x_1\cdots\Delta x_d}\int_{I_{j_1,\dots,j_d}} u\,dx, \quad x=(x_1,\dots,x_d)\in I_{j_1,\dots,j_d}, \tag{4.12}$$

and a projection P_δ of the cell averages onto the values in V_δ. Furthermore, we let f_i^δ, $i=1,\dots,d$, denote the continuous piecewise linear approximations to f_i, $i=1,\dots,d$, as in (4.2). We set $\eta=(\delta,\Delta x_1,\dots,\Delta x_d,\Delta t)$ and let u^0 denote the projection of u_0 on the grid, that is $u^0=\pi_\delta u_0$. Let $S^{f_i^\delta,x_i}(t)$ denote the solution operator of the scalar conservation law in one dimension, viz.,

$$\begin{aligned}(v_i^\delta)_t + f_i^\delta(v_i^\delta)_{x_i} &= 0, \quad (x_i,t)\in\mathbb{R}\times(0,T),\\ v_i^\delta(x_i,0) &= v_{i0}^\delta(x_i), \quad x_i\in\mathbb{R},\end{aligned}$$

that is, we write $v(x_i,t)=S^{f_i^\delta,x_i}(t)v_{i0}^\delta$. Since v_{i0}^δ is piecewise constant, and f_i^δ piecewise linear, the solution can be calculated using front tracking.

Then we obtain an approximation of the solution to (3.1) – (3.3) by successively applying the front tracking solution operator $S^{f_i^\delta,x_i}(t)$ followed by the projection operator π_δ (in order to prevent the number of discontinuities from growing excessively). We denote the approximate solutions at the time steps $t_r=r\Delta t$, $t\in\mathbb{Q}$ by

$$u^{n+i/d} = \pi_\delta\circ S^{f_i^\delta,x_i}(\Delta t)\,u^{n+(i-1)/d}, \quad i=1,\dots,d,\ n\in\mathbb{N},$$

and

$$u^\eta(x,t) = \begin{cases} S^{f_i^\delta,x_i}(d(t-t_{n+(i-1)/d}))\,u^{n+(i-1)/d}, & t\in[t_{n+(i-1)/d},t_{n+i/d}),\\ u^{n+i/d}, & t=t_{n+i/d},\end{cases} \tag{4.13}$$

$i=1,\dots,d$ and $n\in\mathbb{N}$. The approximation u^η satisfies (see [**73**, Chapter 4]):

THEOREM 4.8. *Let $u_0\in BV(\mathbb{R}^d)\cap L^\infty(\mathbb{R}^d)$ and $f_i(u)\in\mathrm{Lip}(\mathbb{R})$ and piecewise C^2. Then the function u^η defined in* (4.13) *satisfies*

(i) Uniform bound in $\eta = (\delta, \Delta x_1, \ldots, \Delta x_d, \Delta t)$ *for all* $t \in (0, \infty)$*:*

$$\|u^\eta(\cdot, t)\|_{L^\infty(\mathbb{R}^d)} \le \|u_0\|_{L^\infty(\mathbb{R}^d)}, \quad t \in (0, \infty),$$

(ii) The total variation of u^η *is bounded by the total variation of the initial data for all times* $t \in (0, \infty)$*,*

$$\mathrm{TV}(u^\eta(\cdot, t)) \le \mathrm{TV}(u_0), \quad t \in (0, \infty),$$

(iii) For any sequence $\{\eta_j\}_{j\in\mathbb{N}}$*, where* $\eta_j \to 0$ *when* $j \to \infty$*, satisfying*

$$\max_{i=1,\ldots,d} \Delta x_i/\Delta t \le K < \infty,$$

the corresponding sequence $\{u^{\eta_j}\}_{j\in\mathbb{N}}$ *converges in* $C([0,T]; L^1_{\mathrm{loc}}(\mathbb{R}^d))$ *to the unique entropy solution* u *of* (3.1)–(3.3)*. Specifically, we have, denoting* $\|f\|_{\mathrm{Lip}} = \max_{i=1,\ldots,d} \|f_i\|_{\mathrm{Lip}(\mathbb{R})}$ *and* $\Delta x = \max_{i=1,\ldots,d} \Delta x_i$*,*

$$\begin{aligned}(4.14)\quad &\|u(\cdot,t) - u^\eta(\cdot,t)\|_{L^1(\mathbb{R}^d)} \\ &\le \|u_0 - u^0\|_{L^1(\mathbb{R}^d)} + t\|f - f^\delta\|_{\mathrm{Lip}(\mathbb{R})}\,\mathrm{TV}(u_0) \\ &\quad + 2\,\mathrm{TV}(u_0)\,\sqrt{2t}\,(\sqrt{d}+1)\sqrt{d\Delta x^2/\Delta t + \Delta x\|f\|_{\mathrm{Lip}} + \Delta t\|f\|^2_{Lip}}\,.\end{aligned}$$

Corollary 4.9. *Under the assumptions of Theorem* 4.8 *and choosing the parameters* Δx*,* Δt *and* δ *as*

$$\Delta x = k_1 \Delta t = k_2 \delta^2, \tag{4.15}$$

where k_1 *and* k_2 *are positive constants, the dimensional splitting front tracking algorithm converges at rate* 1 *in the parameter* δ*, specifically,*

$$\|u(\cdot,t) - u^\eta(\cdot,t)\|_{L^1(\mathbb{R}^d)} \le C\,\delta\,(1+t)\Big(1 + \|f\|_{W^{2,\infty}(\mathbb{R};\mathbb{R}^d)}\Big)\,\mathrm{TV}(u_0), \tag{4.16}$$

where $C > 0$ *is a constant depending at most linearly on* d*.*

Proof. Similarly to Corollary 4.3, we have that the approximation u^0 of the initial data u_0 satisfies

$$\|u_0 - u^0\|_{L^1(\mathbb{R}^d)} \le c\,d\,\delta\,\mathrm{TV}(u_0),$$

and (4.8), (4.14) yields a convergence rate with respect to the parameters Δx, Δt and δ,

$$\begin{aligned}\|u(\cdot,t) - u^\eta(\cdot,t)\|_{L^1(\mathbb{R}^d)} \le \Big(&c\,d\,\delta + t\,\delta\,\|f\|_{W^{2,\infty}(\mathbb{R};\mathbb{R}^d)} \\ &+ 2\sqrt{2t}\,(\sqrt{d}+1)\sqrt{d\Delta x^2/\Delta t + \Delta x\|f\|_{\mathrm{Lip}} + \Delta t\|f\|^2_{\mathrm{Lip}}}\Big)\,\mathrm{TV}(u_0).\end{aligned}$$

We see that this yields 4.16 if we choose Δx, Δt and δ as in (4.15). □

We next estimate the convergence rate of the dimensional splitting front tracking algorithm with respect to the work needed to compute one approximation of the solution.

LEMMA 4.10. *(Work estimate for $d \geq 2$) Under the assumptions of Theorem* 4.8 *and* (4.15)*, the front tracking approximation satisfies,*

$$\|u(\cdot,t) - u^\eta(\cdot,t)\|_{L^1(\mathbb{R}^d)} \leq C\,(1+t^{(2d+3)/(2d+2)})\Big(1+\|f\|_{W^{2,\infty}}\Big) \times \mathrm{TV}(u_0)\left((\|u_0\|_{L^\infty}+1)\,(\|u_0\|_{L^\infty}+\mathrm{TV}(u_0))\right)^{1/(2(d+1))} (W_{\delta,d}^{FT})^{-1/(2(d+1))}\,, \tag{4.17}$$

where $C>0$ is a constant depending only on d.

PROOF. The work done in one time interval $(t_{n+(i-1)/d}, t_{n+i/d}]$ consists of two components, the front tracking approximation in $(t_{n+(i-1)/d}, t_{n+i/d})$ and the projections at time $t = t_{n+i/d}$. As in the one-dimensional case, we can solve all possible Riemann problems beforehand and store the solutions, the work to do this is of order $C\,\overline{R}^2\,d\,\delta^{-2}$, where $\overline{R} = \|u_0\|_{L^\infty}$, since the flux f has d components f_i (see Remark 4.5). Then the work for the front tracking approximation in $(t_{n+(i-1)/d}, t_{n+i/d})$ is of the order of the number of interactions of fronts $T(\eta,t)$ in that time interval. This number is bounded by

$$T(\eta,t) \leq C\,(\|u_0\|_{L^\infty}+1)\,(\mathrm{TV}(u_0)+\|u_0\|_{L^\infty})\,\delta^{-2}(\Delta x)^{-(d-1)},$$

which is (4.10) multiplied by $(\Delta x)^{-(d-1)}$, because we do the front tracking in each segment $I^i_{j_1,\dots,j_d} := [j_1\Delta x,(j_1+1)\Delta x)\times\cdots\times[j_{i-1}\Delta x,(j_{i-1}+1)\Delta x)\times\mathbb{R}\times\cdots\times[j_d\Delta x,(j_d+1)\Delta x)$. The work $W^{\pi_\delta}_{t_{n+i/d}}$ needed to do the projections at time $t_{n+i/d}$ is of the same order,

$$W^{\pi_\delta}_{t_{n+i/d}} = C(\|u_0\|_{L^\infty(\mathbb{R})}+1)\,(\mathrm{TV}(u_0)+\|u_0\|_{L^\infty})\,\delta^{-2}(\Delta x)^{-(d-1)},$$

as it is proportional to the number of fronts in the x_i-direction and the number of segments $I^i_{j_1,\dots,j_d}$. Hence the total work $W^{FT}_{\delta,d}$ needed to compute the front tracking approximation $u^\eta(\cdot,t)$ is of order

$$W^{FT}_{\delta,d} = C\,t\,d\,(\|u_0\|_{L^\infty(\mathbb{R})}+1)\,(\mathrm{TV}(u_0)+\|u_0\|_{L^\infty})\,\delta^{-2}(\Delta x)^{-(d-1)}(\Delta t)^{-1}\,.$$

Now using (4.15), we obtain the convergence estimate with respect to work, (4.17). □

REMARK 4.11. Observe that the convergence rate (4.17) is of the same order with respect to the work $W^{FT}_{\delta,d}$ as the one for the approximation by a finite volume scheme (see e.g. [**97**]). So in contrast to the one-dimensional case we do not get an improvement of the rate by using the front tracking method.

REMARK 4.12 (Work estimate for convex flux functions). As in the case $d=1$, the estimate on the total work $W^{FT}_{\delta,d}$ can be improved if the components f_i, $i=1,\dots,d$ of the flux function are convex. Again, solving a Riemann problem with left state u_L and right state u_R reduces to checking whether $u_L > u_R$. Moreover, the total number of interactions in each time interval $t\in(t_{n+(i-1)/d}, t_{n+i/d})$ is bounded by $T(\eta,t)\leq \mathrm{TV}(u_0)\delta^{-1}$ and therefore,

$$\|u(\cdot,t) - u^\eta(\cdot,t)\|_{L^1(\mathbb{R}^d)} \leq C\left(1+t^{(2d+2)/(2d+1)}\right) \times\Big(1+\|f\|_{W^{2,\infty}}\Big)\,\mathrm{TV}(u_0)^{(2d+2)/(2d+1)}\,\left(W^{\mathrm{FT}}_{\delta,d}\right)^{-1/(2d+1)},$$

for convex or concave flux functions.

4.2.3. *Front tracking for RSCLs.* Having described the convergence properties of the front tracking algorithm for deterministic scalar conservation laws, we are ready to state the convergence result for the approximation of the random scalar conservation law (3.14):

THEOREM 4.13. *Assume that the random (as in Definition* 3.2*) initial data* u_0 *and flux function* f *satisfy* (3.12).

For $\delta > 0$, *let* $f_i^\delta(\omega, \cdot)$ *denote the piecewise linear interpolations to the random flux component functions* $f_i(\omega, \cdot)$ *as defined in* (4.2).

Let the discretization parameter vector $\eta = \delta$ *if* $d = 1$, *and* $\eta = (\delta, \Delta x_1, \dots, \Delta x_d, \Delta t)$ *if* $d > 1$, *and let* $u^\eta(\omega; \cdot, \cdot)$ *denote the corresponding approximate solution defined by* (4.3a) *if* $d = 1$ *and* (4.13) *if* $d > 1$, *with initial data* $u_0(\omega; \cdot)$ *and flux functions* $f_1(\omega; \cdot), \dots, f_d(\omega; \cdot)$. *Then the approximations* u^η *satisfy*

$$\|u^\eta(\cdot; \cdot, t)\|_{L^\infty(\Omega; L^\infty(\mathbb{R}^d))} \le \overline{M}, \quad t \in (0, \infty),$$

the total variation is bounded $\mathbb{P}$*-almost surely,*

$$\mathrm{TV}(u^\eta(\omega; \cdot, t)) \le \mathrm{TV}(u_0(\omega; \cdot)), \quad t \in (0, \infty),\ \mathbb{P}\text{-a.e. } \omega \in \Omega\,.$$

As $\eta \to 0$, *the sequence* $(u^\eta)_{\eta > 0}$ *converges* $\mathbb{P}$*-almost surely and in* $C([0, T]; L^1(\mathbb{R}^d))$, *to the unique random entropy solution of the RSCL* (3.14). *Moreover, if* $d = 1$, *we have* $\mathbb{P}$*-a.s. the error bound*

$$\begin{aligned}&\|u(\omega; \cdot, t) - u^\eta(\omega; \cdot, t)\|_{L^1(\mathbb{R})}\\ &\qquad \le \|u_0(\omega; \cdot) - u^0(\omega; \cdot)\|_{L^1(\mathbb{R})} + t\|f(\omega; \cdot) - f^\delta(\omega; \cdot)\|_{\mathrm{Lip}(\mathbb{R})}\,\mathrm{TV}(u_0(\omega; \cdot)),\end{aligned}$$

and if $d > 1$, *we have* $\mathbb{P}$*-a.s.*

$$\begin{aligned}(4.18)\quad &\|u(\omega; \cdot, t) - u^\eta(\omega; \cdot, t)\|_{L^1(\mathbb{R}^d)}\\ &\le \|u_0(\omega; \cdot) - u^0(\omega; \cdot)\|_{L^1(\mathbb{R}^d)} + t \max_{i=1,\dots,d} \|f_i(\omega; \cdot) - f_i^\delta(\omega; \cdot)\|_{\mathrm{Lip}(\mathbb{R})}\,\mathrm{TV}(u_0(\omega; \cdot))\\ &\quad + 2\ \mathrm{TV}(u_0(\omega; \cdot))\,\sqrt{2t}\,(\sqrt{d} + 1)\sqrt{d\Delta x^2/\Delta t + \Delta x\|f(\omega; \cdot)\|_{\mathrm{Lip}} + \Delta t\|f(\omega; \cdot)\|_{Lip}^2}.\end{aligned}$$

PROOF. The assertion follows from Theorems 4.2 and 4.8 upon noting that the assumptions given there are satisfied pathwise, i.e., for $\mathbb{P}$-a.e. $\omega \in \Omega$. □

From now on we assume that

$$f(\omega; \cdot) \in L^\infty(\Omega; W^{2,\infty}([-\overline{M}, \overline{M}]; \mathbb{R}^d)) \tag{4.19}$$

where $\overline{M}$ is as in (3.12).

COROLLARY 4.14. *Under the assumption* (4.19), *choose* $\Delta x = k_1\delta$ *for* $d = 1$ *and* $\Delta x = k_1\Delta t = k_2\delta^2$ *for* $d \ge 2$. *Then*

$$\begin{aligned}(4.20)\quad &\|u(\omega; \cdot, t) - u^\eta(\omega; \cdot, t)\|_{L^1(\mathbb{R}^d)}\\ &\qquad \le C\,\delta\,(1 + t)\left(1 + \|f(\omega; \cdot)\|_{W^{2,\infty}([-\overline{M}, \overline{M}]; \mathbb{R}^d)}\right)\mathrm{TV}(u_0(\omega; \cdot))\,.\end{aligned}$$

If in addition $u_0 \in L^p(\Omega; BV(\mathbb{R}^d))$ *and* $f \in L^q(\Omega; W^{2,\infty}(\mathbb{R}; \mathbb{R}^d))$ *for some* $1 \le p, q \le \infty$ *with* $1/p + 1/q = 1$*, we have*

$$\begin{aligned} \|\mathbb{E}[u(t)] - \mathbb{E}[u^\eta(t)]\|_{L^1(\mathbb{R}^d)} &\le \|u(t) - u^\eta(t)\|_{L^1(\Omega; L^1(\mathbb{R}^d))} \\ &\le C\,\delta\,(1+t)\Big(1 + \|f\|_{L^q(\Omega; W^{2,\infty})}\Big)\,\|\operatorname{TV}(u_0)\|_{L^p(\Omega)}, \end{aligned} \tag{4.21}$$

for all δ *and* $t > 0$*.*

Proof. The bound (4.20) follows from the regularity assumption on $f(\omega, \cdot)$, and the inequality (4.21) is proved by an application of Hölder's inequality to (4.20), and by using (2.2). □

4.2.4. *Multilevel flux decomposition.* The approximate, continuous, piecewise linear flux functions f_i^δ defined by (4.2) are particular useful in connection with empirical flux data (such as typically arise in Buckley-Leverett models where flux functions are built from empirical data) and with MLMC, as will be seen in the next subsection.

We choose $\delta_0 > 0$ and let $\delta_\ell = 2^{-\ell}\delta_0$. Let also $f_i^\ell(\omega; \cdot) := f_i^{\delta_\ell}(\omega; \cdot)$ denote the continuous piecewise linear interpolant of $f_i(\omega; \cdot)$, for $i = 1, \dots, d$, as defined by (4.2), and similarly set $f^\ell := (f_1^\ell, \dots, f_d^\ell)$.

Lemma 4.15. *Under assumption* (4.19)*, for* $\ell = 0, 1, 2, ...$*, the continuous, piecewise linear flux interpolants* $f_i^\ell(\omega; \cdot) = f_i^{\delta_\ell}(\omega; \cdot)$ *are bounded random flux functions in the sense of Definition* 3.2 *which satisfy the bound* (3.12) *with constant* $\overline{M}$ *which is independent of* ℓ*, and which satisfy for* $\mathbb{P}$*-a.e.* $\omega \in \Omega$ *the error bound*

$$\|f_i(\omega; \cdot) - f_i^\ell(\omega; \cdot)\|_{W^{1,\infty}([-\overline{M}, \overline{M}]; \mathbb{R}^d)} \le C 2^{-\ell} \|\partial_u^2 f_i(\omega; \cdot)\|_{L^\infty([-\overline{M}, \overline{M}])} \tag{4.22}$$

Proof. The proof of (4.22) follows from standard approximation estimates for the nodal interpolation. □

The following corollary is a direct consequence of (4.22).

Corollary 4.16. *Under the assumptions of Lemma* 4.15*, we have*

$$\|(f_i^\ell - f_i^{\ell-1})(\omega; \cdot)\|_{\mathrm{Lip}([-\overline{M}, \overline{M}]; \mathbb{R}^d)} \le 2C 2^{-\ell} \|\partial_u^2 f_i(\omega; \cdot)\|_{L^\infty([-\overline{M}, \overline{M}]; \mathbb{R})}\,.$$

Here, the constant $C > 0$ *is independent of* ℓ *and of the flux* f*.*

4.3. MLMC front tracking. The MLMC discretization of differential equations with random inputs was proposed by Giles in [**54, 55**], upon earlier work by Heinrich on numerical integration in [**64**]. For random scalar conservation laws (RSCLs), the MLMC Finite Volume discretizations were proposed and analyzed, in the case of deterministic flux and random initial conditions, in [**97**], and for RSCLs with random flux, in [**96**].

Here, we analyze the convergence of MLMC in conjunction with Front Tracking (FT) discretizations. Although the analysis proceeds, broadly speaking, along the lines of what was done in [**97, 96**], there are notable differences: First, unlike [**96**], there is no need for a principal component analysis of the random flux, e.g. via a Karhunen–Loève expansion. Secondly, we propose the use of a *multiresolution decomposition of the random flux* on the

phase space of the solution. Finally, the error bounds which we shall obtain relate, in a rather explicit fashion, the number M_ℓ of MC samples on different discretization levels to the flux variance at resolution ℓ, i.e., to $\|f^\ell - f^{\ell-1}\|^2_{L^2(\Omega;\mathrm{Lip}(\mathbb{R},\mathbb{R}^d))}$. Since f^ℓ is piecewise linear, this quantity can easily be computed for empirically calibrated random flux functions and, thereby, the number M_ℓ of "samples" (which are approximate solutions of the RSCL with flux functions f^ℓ and $f^{\ell-1}$, obtained by front tracking), can be scaled accordingly.

We start the analysis by introducing some notation. For $d = 1$, we let $\Delta x_\ell = \delta_l = 2^{-\ell}\delta_0$ for some $\delta_0 > 0$. For $d \geq 2$, $\ell = 0, 1, 2, \ldots$, we set

$$\eta_\ell = (\delta_\ell, \Delta x_\ell, \Delta t_\ell) = \left(2^{-\ell}\delta_0, 2^{-2\ell}\Delta x_0, 2^{-2\ell}\Delta t_0\right).$$

Moreover, we let $u_0^\ell(\omega;\cdot) := \pi_\ell u_0(\omega;\cdot)$ where $\pi_\ell = P_{\Delta x_\ell} \circ \overline{P}_{\Delta x_\ell}$, cf. (4.12). Note that we set $\Delta x_1 = \cdots = \Delta x_d = \Delta x_\ell$.

Then we denote for $\ell = 0, 1, 2, ...$, by $u^\ell(\omega; x, t)$ the approximations of $u(\omega; x, t)$ obtained by the front tracking method with initial data u_0^ℓ and f^ℓ.

As in [**97**], $E_M[\cdot]$ denotes the sample average of M i.i.d. samples of a random quantity. We are interested in the computation of the statistical mean

$$\mathbb{E}[u(t)] \in C([0,T]; L^1(\mathbb{R}^d))$$

of the random entropy solution of the RSCL (3.1) - (3.3). To this end, the MLMC-FT approximation is defined as follows: for a given level $L \in \mathbb{N}$ of refinement, we use the linearity of the mathematical expectation $\mathbb{E}[\cdot]$ to write

$$\mathbb{E}[u(t)] \simeq \mathbb{E}[u^L(t)] = \sum_{\ell=0}^{L} \mathbb{E}\left[u^\ell - u^{\ell-1}\right].$$

Here, and in the following, we adopt the convention that $u^{-1} \equiv 0$.

We next *estimate the expectations of increments for each level of refinement by a level-dependent number M_ℓ of samples*, which results in the MLMC estimate

$$E_L^{MLMC}[u^L(t)] := \sum_{\ell=0}^{L} E_{M_\ell}\left[u^\ell - u^{\ell-1}\right]. \tag{4.23}$$

Here, u^ℓ are the approximations obtained by front tracking for the initial data u_0^ℓ and the flux functions f^ℓ.

4.4. Convergence analysis. We are now interested in estimating

$$\mathbb{E}[u(t)] - E_L^{MLMC}[u^L(t)].$$

To this end, we write

$$\mathbb{E}[u(t)] - E_L^{MLMC}[u^L(t)] = \underbrace{\mathbb{E}[u(t)] - \mathbb{E}[u^L(t)]}_{A} + \underbrace{\mathbb{E}[u^L(t)] - E_L^{MLMC}[u^L(t)]}_{B}.$$

We have already estimated the $L^1(\mathbb{R}^d)$-norm of term A in equation (4.21). In this setting, it is of order $\mathcal{O}(2^{-L})$ under the additional assumption that $u_0 \in L^p(\Omega; BV(\mathbb{R}^d))$ and $f \in$

$L^q(\Omega; W^{2,\infty}(\mathbb{R};\mathbb{R}^d))$, where $1/p + 1/q = 1$. Consider now the term B. To estimate it, we write, with $\Delta u^\ell := u^\ell - u^{\ell-1}$ for $\ell = 0, 1, 2, ..., L$ and with the convention that $u^{-1} \equiv 0$,

$$\begin{aligned}\|\mathbb{E}[u^L(t)] - E_L^{MLMC}[u^L(t)]\|^2_{L^2(\Omega;L^1(\mathbb{R}^d))} &\\ &= \Big\|\mathbb{E}\Big[\sum_{\ell=0}^{L}(u^\ell - u^{\ell-1})\Big] - E_L^{MLMC}[u^L(t)]\Big\|^2_{L^2(\Omega;L^1(\mathbb{R}^d))}\\ &= \Big\|\sum_{\ell=0}^{L}\big\{\mathbb{E}[\Delta u^\ell] - E_{M_\ell}[\Delta u^\ell]\big\}\Big\|^2_{L^2(\Omega;L^1(\mathbb{R}^d))}.\end{aligned}$$

Expanding the square, and interpreting the M_ℓ samples as i.i.d. copies of the random variable $u^\ell(\omega; x, t)$, we obtain

$$\|\mathbb{E}[u^L(t)] - E_L^{MLMC}[u^L(t)]\|^2_{L^2(\Omega;L^1(\mathbb{R}^d))} = \sum_{\ell=0}^{L}\big\|\mathbb{E}[\Delta u^\ell] - E_{M_\ell}[\Delta u^\ell]\big\|^2_{L^2(\Omega;L^1(\mathbb{R}^d))}.$$

Next we estimate each term in the sum as follows:

$$\begin{aligned}B_\ell &:= \big\|\mathbb{E}[\Delta u^\ell] - E_{M_\ell}[\Delta u^\ell]\big\|^2_{L^2(\Omega;L^1(\mathbb{R}^d))}\\ &= \frac{1}{M_\ell}\mathbb{E}\left[\|\mathbb{E}[\Delta u^\ell(t)] - \Delta u^\ell(t)\|^2_{L^1(\mathbb{R}^d)}\right]\\ &\le \frac{1}{M_\ell}\|\Delta u^\ell(t)\|^2_{L^2(\Omega;L^1(\mathbb{R}^d))}.\end{aligned}$$

We use the elementary estimate

$$\|\Delta u^\ell(\omega;\cdot,t)\|^2_{L^1(\mathbb{R}^d)} \le 2\|u(\omega;\cdot,t) - u^\ell(\omega;\cdot,t)\|^2_{L^1(\mathbb{R}^d)} + 2\|u(\omega;\cdot,t) - u^{\ell-1}(\omega;\cdot,t)\|^2_{L^1(\mathbb{R}^d)}$$

and the convergence rate (4.20), to obtain

$$\|u(t) - u^\ell(t)\|_{L^2(\Omega;L^2(\mathbb{R}^d))} \le C\,\delta_\ell\,(1+t)\Big(1 + \|f\|_{L^2(\Omega;W^{2,\infty})}\Big)\,\|\operatorname{TV}(u_0)\|_{L^\infty(\Omega)}.$$

under the assumption that $u_0 \in L^\infty(\Omega; BV(\mathbb{R}^d))$ and $f \in L^2(\Omega; W^{2,\infty}(\mathbb{R};\mathbb{R}^d))$. Thus,

$$B_\ell \le \frac{1}{M_\ell} C\,\delta_\ell^2\,(1+t^2)\Big(1 + \|f\|^2_{L^2(\Omega;W^{2,\infty})}\Big)\,\|\operatorname{TV}(u_0)\|^2_{L^\infty(\Omega)},$$

where $C > 0$ is a constant which depends on d but which is independent of t. Summing over $\ell = 0, \ldots, L$, we arrive at

$$\begin{aligned}\|\mathbb{E}[u^L(t)] - E_L^{MLMC}[u^L(t)]\|^2_{L^2(\Omega;L^1(\mathbb{R}))} &\\ &\le C\,(1+t^2)\sum_{\ell=0}^{L}\frac{1}{M_\ell}\,\delta_\ell^2\,\Big(1 + \|f\|^2_{L^2(\Omega;W^{2,\infty})}\Big)\,\|\operatorname{TV}(u_0)\|^2_{L^\infty(\Omega)}.\end{aligned}$$

We can now state our basic MLMC-FT error bound.

THEOREM 4.17. *Consider the RSCL with random data* (u_0, f) (3.11) *in the sense of Definition* 3.2 *and satisfying* (3.12). *Assume for* $\overline{M}$ *as in* (3.12) *that* (4.19) *holds.*

Then, for any $L \in \mathbb{N}$ *and for any choice of samples sizes* $\{M_\ell\}_{\ell=0}^L$ *in the MLMC-FT estimator* $E_L^{MLMC}[u^L(t)]$ *in* (4.23) *we have the error bound*

$$\begin{aligned}
&\left\|\mathbb{E}[u(t)] - E_L^{MLMC}[u^L(t)]\right\|^2_{L^2(\Omega;L^1(\mathbb{R}^d))} \\
&\quad\le 2\,C(1+t^2)\delta_L^2\left(1+\|f\|^2_{L^1(\Omega;W^{2,\infty})}\right)\|\operatorname{TV}(u_0)\|^2_{L^\infty(\Omega)} \\
&\qquad + C\,(1+t^2)\sum_{\ell=0}^{L}\frac{1}{M_\ell}\,\delta_\ell^2\,\left(1+\|f\|^2_{L^2(\Omega;W^{2,\infty})}\right)\|\operatorname{TV}(u_0)\|^2_{L^\infty(\Omega)} \\
&\quad\le C\Big[2^{-2L}+\sum_{\ell=0}^{L}M_\ell^{-1}2^{-2\ell}\Big](1+t^2) \\
&\qquad\qquad \times\left(1+\|f\|^2_{L^2(\Omega;W^{2,\infty})}\right)\|\mathrm{TV}(u_0)\|^2_{L^\infty(\Omega)}\,.
\end{aligned}$$

With the particular choice

$$M_\ell = 2^{2(L-\ell)}\,,\quad \ell = 0,\dots,L,$$

we find for any $0 \le t \le T < \infty$ *the bound*

$$\begin{aligned}
&\left\|\mathbb{E}[u(t)] - E_L^{MLMC}[u^L(t)]\right\|^2_{L^2(\Omega;L^1(\mathbb{R}^d))} \\
&\qquad\qquad \le C\,L2^{-2L}\,(1+t^2)\Big(1+\|f\|^2_{L^2(\Omega;W^{2,\infty})}\Big)\|\operatorname{TV}(u_0)\|^2_{L^\infty(\Omega)}.
\end{aligned}\tag{4.24}$$

PROOF. The proof follows from the foregoing analysis. □

If we denote the work for one FT solution at mesh level ℓ by W_ℓ^{FT}, and use the front tracking work estimates in Lemmas 4.5 and 4.10, we obtain the work estimate $W_{L,MLMC}^{FT}$ for the MLMC front tracking method,

$$W_{L,MLMC}^{FT} = C\sum_{\ell=0}^{L}M_\ell W_\ell^{FT} = \begin{cases}\mathcal{O}(W_L^{FT}\,\log W_L^{FT}) = \mathcal{O}(L\,\delta_L^{-2}) & \text{if } d=1,\\ \mathcal{O}(W_L^{FT}) = \mathcal{O}(\delta_L^{-2(d+1)}) & \text{if } d\ge 2.\end{cases}\tag{4.25}$$

This gives us the convergence rates for the MLMC-FT estimator $E_L^{MLMC}[u^L(t))]$ with respect to work:

COROLLARY 4.18. *Under the assumptions of Theorem* 4.17, *the MLMC-FT estimator* $E_L^{MLMC}[u^L(t))]$ *converges with the following rates to the ensemble average* $\mathbb{E}[u(t)]$ *of the random entropy solution*

$$\left\|\mathbb{E}[u(t)] - E_L^{MLMC}[u^L(t)]\right\|^2_{L^2(\Omega;L^1(\mathbb{R}))} \le C\,(\log W_{L,MLMC}^{FT})^2\,(W_{L,MLMC}^{FT})^{-1},\tag{4.26}$$

for $d = 1$, *and*

$$\left\|\mathbb{E}[u(t)] - E_L^{MLMC}[u^L(t)]\right\|^2_{L^2(\Omega;L^1(\mathbb{R}^d))} \le C\,(\log W_{L,MLMC}^{FT})\,(W_{L,MLMC}^{FT})^{-1/(d+1)}$$

for $d \geq 2$, where $C > 0$ is a constant depending on d and t, and on $\|u_0\|_{L^\infty(\Omega;BV(\mathbb{R}^d))}$ and $\|f\|_{L^2(\Omega;W^{2,\infty}(-\overline{M},\overline{M};\mathbb{R}^d))}$.

REMARK 4.19. We have seen in Lemma 4.7 that the convergence rate of the deterministic front tracking algorithm for $d = 1$ is one with respect to work, if the flux function f is convex. However, this does not show up as an improvement of the convergence rate of the MLMC-FT method, since in this case the work of the Monte Carlo method dominates. Specifically, in the case of a convex flux and $d = 1$, we have

$$\begin{aligned} W^{FT}_{L,MLMC} &= C \sum_{\ell=0}^{L} M_\ell \, W^{FT}_\ell \leq C \sum_{\ell=0}^{L} M_\ell \, \delta_\ell^{-1} \\ &\leq C \, 2^{2L} \sum_{\ell=0}^{L} 2^{-2\ell} \, 2^{\ell} \leq C \, 2^{2L} = \mathcal{O}(\delta_L^{-2}), \end{aligned} \tag{4.27}$$

which is the same effort as in the general case (4.25) apart from the missing factor L.

This is to be contrasted to several space dimensions, where we have a small gain in convergence rate if all the flux components f_j, $j = 1, \ldots, d$ are convex, since the convergence rate of the deterministic dimensional splitting front tracking method is worse than that of the Monte Carlo method:

$$\begin{aligned} W^{FT}_{L,MLMC} &= C \sum_{\ell=0}^{L} M_\ell \, W^{FT}_\ell \leq C \sum_{\ell=0}^{L} M_\ell \, \delta_\ell^{-(2d+1)} \\ &\leq C \, 2^{2L} \sum_{\ell=0}^{L} 2^{(-1+2d)\ell} \leq C \, 2^{(2d+1)L} = \mathcal{O}(\delta_L^{-(2d+1)}). \end{aligned}$$

5. Numerical experiments

In this section, we test the performance of the MLMC-FT method on several examples with random fluxes in one and two space dimensions.

5.1. Convex random flux in one space dimension. We consider the random scalar conservation law,

$$u_t + f(\omega; u)_x = 0, \quad x \in [-1, 1], \, t \in (0, \infty), \tag{5.1a}$$

$$u(\omega; x, 0) = -\sin(\pi x), \quad x \in [-1, 1], \, t = 0, \tag{5.1b}$$

with periodic boundary conditions and the random flux $f(\omega; u)$ given by

$$f(\omega; u) = \frac{1}{p(\omega)} |u|^{p(\omega)}, \quad p(\omega) \sim \mathcal{U}(1.5, 2.5). \tag{5.2}$$

This flux function is a bounded random flux and for $\mathbb{P}$-a.e., $f(\omega; \cdot) \in \mathrm{Lip}([-\overline{M}, \overline{M}]; \mathbb{R})$, where $\overline{M} \geq \|u_0\|_{L^\infty(\mathbb{R})}$ is as in (3.12). An approximation of the mean of the random entropy solution at time $t = 1$, computed by the MLMC-FT method for $L = 9$, with $\delta_0 = 2^{-4}$ at the coarsest level, and $M_L = 8$ samples at the level with the finest resolution, is shown in Figure 1. In order to compute an estimate on the error of the approximation

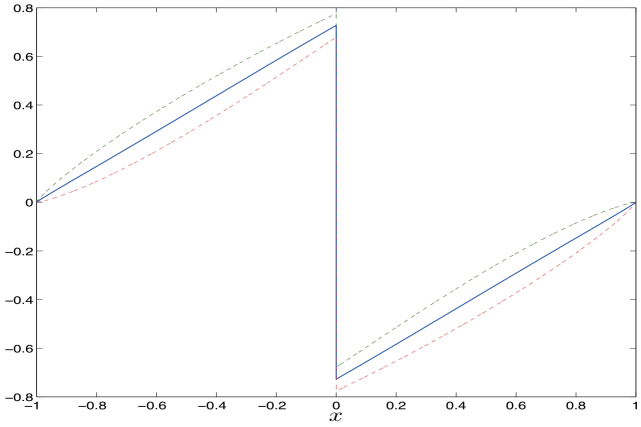

FIGURE 1. The estimator $E_L^{MLMC}[u^L(t)]$ computed by the MLMCFT method at time $t = 1$ with $L = 9$ for problem (5.1), (5.2). The dashed lines denote the mean with $\pm$ standard deviation.

of the mean by the MLMC estimator $E_L^{MLMC}[u^L(t)]$ in the $L^2(\Omega; L^1(\mathbb{R}))$-norm, we use the relative error estimator introduced in [**97**] based on a Monte Carlo quadrature in the stochastic domain: We denote by U_{ref} a reference solution and by $\{U_k\}_{k=1,\dots,K}$ a sequence of statistically independent approximate solutions $E_L^{MLMC}[u^L(t)]$ obtained by running the MLMC-FT solver K times and corresponding to K realizations in the stochastic domain. Then we estimate the relative error by

$$\mathcal{R}E = \sqrt{\sum_{k=1}^{K} (\mathcal{R}E_k)^2 / K}, \tag{5.3}$$

where

$$\mathcal{R}E_k = 100 \times \frac{\|U_{\text{ref}} - U_k\|_{l^1}}{\|U_{\text{ref}}\|_{l^1}}. \tag{5.4}$$

In [**97**] the sensitivity of the error with respect to the parameter K is investigated. For this example, we will use $K = 30$ which was shown to be sufficient for most problems [**97, 99**]. To compute a reference solution U_{ref}, we have made use of the symmetry properties of the each realization (a shock at $x = 0$, smoothness away from the shock) and used the characteristics of the differential equation to compute an accurate approximation of $\mathbb{E}[u(t)]$. In Figure 2 the errors (5.3) versus the resolution δ_L at the finest level L of the MLMC estimator and versus the run time (in seconds) are shown ($L = 0, \dots, 6$). We observe that the convergence rates are ≈ 0.9 with respect to the resolution and ≈ 0.4

with respect to work, which is approximately what we would expect from the theoretical results: Equation (4.24) implies that the error estimator (5.3) is asymptotically of order $\mathcal{O}(\sqrt{L}\,2^{-L}) = \mathcal{O}(2^{-\alpha(L)\,L}) = \mathcal{O}(\delta_L^{-\alpha(L)})$ with respect to the resolution at the finest level, where

$$\alpha(L) = 1 - \frac{\log L}{2\,L\log 2} \xrightarrow{L\to\infty} 1. \tag{5.5}$$

For $L = 6$, we have $\alpha(L+1) \approx 0.8$. Due to (4.27), the estimator (5.3) is of order $\mathcal{O}((W^{FT}_{L,MLMC})^{-\alpha(L)/2})$ with respect to work, hence for $L = 6$, $\alpha(L+1)/2 \approx 0.4$.

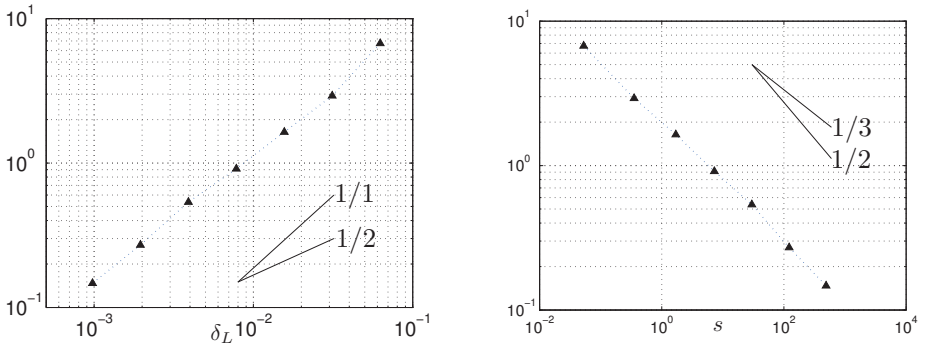

FIGURE 2. Left: Error (5.3) versus the resolution. Right: Error versus the run time of the MLMC-FT solver in seconds for the problem (5.1), (5.2). At the coarsest level, we have used $\delta_0 = 2^{-4}$ and at the finest level, we have used $M_L = 8$ samples.

REMARK 5.1. For exponents $p \in [1.5, 2)$, the second derivative of the flux function $f(u,p)$ in (5.2) is not uniformly bounded. Therefore the bound (4.8) does not apply. However, by a careful refinement of the estimates in [**73**, Chapter 2], it is possible to show that the (deterministic) front tracking method converges at rate one with respect to the discretization parameter δ if the flux function f is in $W^{2,1}([-\overline{M},\overline{M}];\mathbb{R})$ and the initial data $u_0 \in BV(\mathbb{R})$ has a bounded number of local maxima and minima.

5.2. Nonconvex random flux in one space dimension. In a second experiment, we test the performance of the MLMC-FT method on the initial value problem (5.1) with periodic boundary conditions and the nonconvex random flux function

$$f(\omega;u) = \operatorname{sgn}(u)\frac{|u|^{p(\omega)}}{p(\omega)}, \quad p(\omega) \sim \mathcal{U}(2.5, 3.5). \tag{5.6}$$

For $\overline{M} > 0$ as in (3.12), we have $f \in L^2(\Omega; W^{2,\infty}([-\overline{M},\overline{M}];\mathbb{R}))$, hence the assumptions in Theorem 4.17 are satisfied for this problem. In Figure 3, we show an approximation of the

mean of the solution computed by the MLMC-FT-solver at time $t = 1$ with $L = 9$, $\delta_0 = 2^{-5}$ at the coarsest level and $M_L = 4$ samples at the finest level. We see that the mean of the

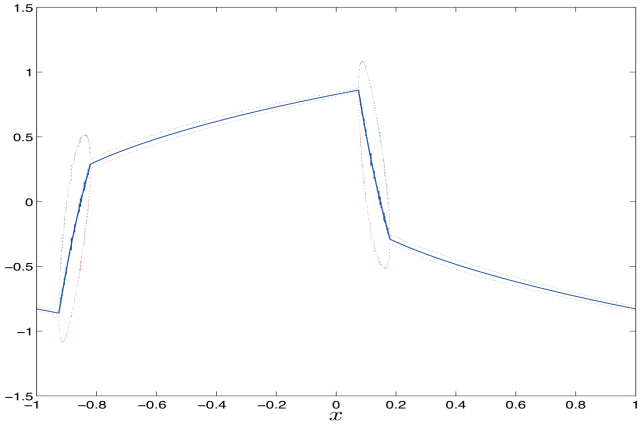

FIGURE 3. The estimator $E_L^{MLMC}[u^L(t)]$ for problem (5.1), (5.6) computed by the MLMC-FT method at time $t = 1$ with $L = 9$. The dashed lines denote the mean with $\pm$ standard deviation.

solution is continuous, whereas all computed pathwise, approximate realizations $u(\omega; \cdot)$ of random entropy solutions of (5.1), (5.6) develop shocks.

This is not unexpected, because while each realization has discontinuities, the location of these discontinuities is random, and disappear upon taking the expectation. However, for each realization, the solution varies (very) rapidly at the shock location, hence the variance will be larger around in the regions where shocks are typically located, than in regions where each realization is continuous. For our example, each realization has two shocks, one around $x = 0.1$ and one around $x = -0.9$. We see that the variance is indeed much larger in around $x = 0.1$ and $x = -0.9$.

We use this approximation as a reference solution and compute the error estimators (5.3), (5.4) for $L = 0, \ldots, 5$, $\delta_0 = 2^{-5}$, $M_L = 4$ and $K = 30$. The results are shown in Figure 4. Similarly to the first example in Section 5.1, the experimentally observed convergence rates validate the a priori estimates (4.24) and (4.26) as we are not yet in the asymptotic regime and for $L = 5$, $\alpha(L + 1) \approx 0.78$, c.f. (5.5) (we observe ≈ 0.85 versus resolution and ≈ 0.35 versus run time).

5.3. Random fluxes in two space dimensions. We test the performance of the MLMC-FT algorithm in several space dimensions on the following test problem,

$$u_t + f(\omega; u)_x + g(\omega; u)_y = 0, \quad (x, y) \in [0, 2]^2,\ t \in (0, \infty), \tag{5.7a}$$

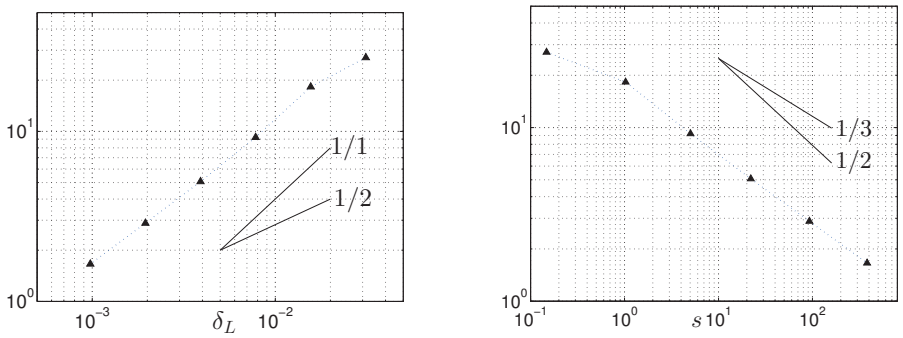

FIGURE 4. Left: Error (5.3) versus the resolution. Right: Error versus the run time of the MLMC-FT solver in seconds for the problem (5.1), (5.6). At the coarsest level, we have used $\delta_0 = 2^{-5}$ and at the finest level, we have used $M_L = 4$ samples.

$$u(\omega; x, y, 0) = \begin{cases} 1, & 0.1 < x, y < 0.9, \\ -1, & (x-1.5)^2 + (y-1.5)^2 < 0.16, \\ 0, & \text{otherwise}, \end{cases} \tag{5.7b}$$

with periodic boundary conditions and random fluxes f and g given by

$$f(\omega; u) = g(\omega; u) = \frac{|u|^{p(\omega)}}{p(\omega)}, \quad p(\omega) \sim \mathcal{U}(1, 3). \tag{5.8}$$

In Section 4.2.2 we have seen that in order to have the optimal convergence rate of the front tracking/dimensional splitting method, we have to choose the grid size Δx, the time step Δt and the refinement parameter δ of the flux function interpolations as

$$\Delta x = k_1 \Delta t = k_2 \delta^2.$$

We call k_1 a *CFL-number* in analogy to finite volume methods, although no restriction needs to be imposed on k_1 since dimensional splitting combined with front tracking method has been shown to converge for any choice of constants $k_1 > 0$.

Due to the increased computational effort of the multidimensional problem compared with the one dimensional problems, we have chosen to refine with respect to the grid size Δx. Therefore we set $\Delta x_\ell = 2^{-\ell} \Delta x_0$ and $\delta_\ell = 2^{-\ell/2} \delta_0$ and use at level $\ell = 0, \dots, L$, $M_\ell = 2^{L-\ell} M_L$ samples. In Figure 5 we show an approximation of the mean of (5.7), (5.8) by the MLMC-FT method computed at time $t = 1$ for $L = 8$ with $M_L = 4$, $\Delta x_0 = 2^{-3}$ and CFL-number $k_1 = 20$. As a reference solution, we use an approximation of the mean of the solution computed by a MLMC-FVM scheme as in [**98**], with an HLL-solver and second order WENO reconstruction, $L = 8$, $M_L = 4$, $\Delta x_0 = 2^{-2}$, on a mesh with $2^{11} \times 2^{11}$ grid cells. We compute the error estimators (5.3), (5.4) for $K = 5$, $L = 0, \dots, 7$,

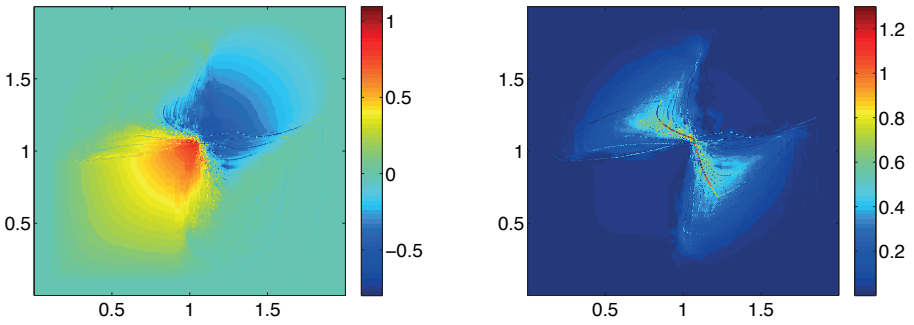

FIGURE 5. Mean and variance of (5.7), (5.8) computed by the MLMC-FT method for $L = 8$, $t = 1$, $M_L = 4$, $\Delta x_0 = 2^{-3}$, CFL-condition $k_1 = 20$ (number of grid cells: $2^{12} \times 2^{12}$). Left: Estimated mean of the solution. Right: Estimated variance of the solution.

$M_L = 4$, $M_\ell = 2^{L-\ell} M_L$, $\Delta x_0 = 0.125$, $\Delta x_\ell = 2^{-\ell} \Delta x_0$. The errors are shown in Figure 6. We measure convergence rates of ≈ 0.45 with respect to the grid size Δx and ≈ 0.15 with respect to the run time of the MLMC-FT solver. From the a priori estimates we would expect rates of $1/2$ versus the grid size and $1/5$ versus work asymptotically, so our rates are slightly below that. This could indicate that we are not yet in the asymptotic regime for the presently considered values of L.

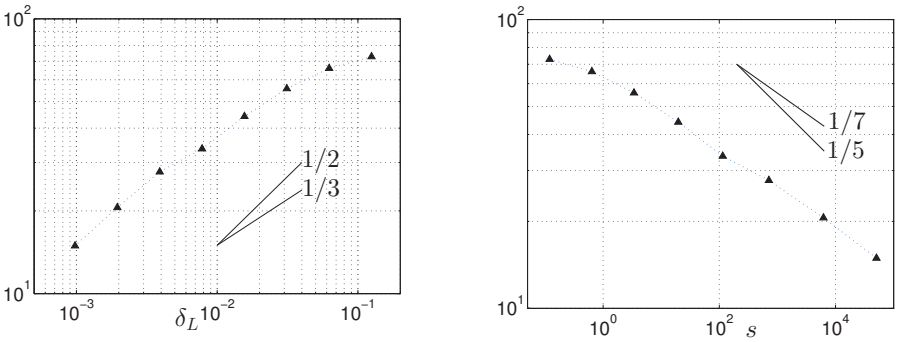

FIGURE 6. Left: Error (5.3) versus the resolution. Right: Error versus the run time of the MLMC-FT solver in seconds (x-axis figure right hand side) for the problem (5.1), (5.2). At the coarsest level, we have used $\Delta x_0 = 2^{-3}$ and at the finest level, we have used $M_L = 4$ samples, $K = 5$.

Bibliography

[1] Adimurthi, S. Mishra, and G. D. V. Gowda. Optimal entropy solutions for conservation laws with discontinuous flux-functions. *J. Hyperbolic Differ. Equ.*, 2(4):783–837, 2005.

[2] S. Agmon. The L_p approach to the Dirichlet problem. I. Regularity theorems. *Ann. Scuola Norm. Sup. Pisa (3)*, 13:405–448, 1959.

[3] B. Andreianov, K. H. Karlsen, and N. H. Risebro. A theory of L^1-dissipative solvers for scalar conservation laws with discontinuous flux. *Arch. Ration. Mech. Anal.*, 201(1):27–86, 2011.

[4] M. A. Austin, P. S. Krishnaprasad, and L. S. Wang. Almost Poisson integration of rigid body systems. *J. Comput. Phys.*, 107(1):105–117, 1993.

[5] K. Aziz and A. Settari. *Petroleum reservoir simulation*. Applied Science Publishers, London, 1979.

[6] K. A. Bagrinovskiĭ and S. K. Godunov. Difference schemes for multidimensional problems. *Dokl. Akad. Nauk SSSR (N.S.)*, 115:431–433, 1957.

[7] Ľ. Baňas, A. Prohl, and R. Schätzle. Finite element approximations of harmonic map heat flows and wave maps into spheres of nonconstant radii. *Numer. Math.*, 115(3):395–432, 2010.

[8] S. Bartels. Semi-implicit approximation of wave maps into smooth or convex surfaces. *SIAM J. Numer. Anal.*, 47(5):3486–3506, 2009.

[9] S. Bartels, X. Feng, and A. Prohl. Finite element approximations of wave maps into spheres. *SIAM J. Numer. Anal.*, 46(1):61–87, 2007/08.

[10] S. Bartels, C. Lubich, and A. Prohl. Convergent discretization of heat and wave map flows to spheres using approximate discrete Lagrange multipliers. *Math. Comp.*, 78(267):1269–1292, 2009.

[11] S. Bartels and A. Prohl. Convergence of an implicit finite element method for the Landau-Lifshitz-Gilbert equation. *SIAM J. Numer. Anal.*, 44(4):1405–1419 (electronic), 2006.

[12] S. Bartels and A. Prohl. Constraint preserving implicit finite element discretization of harmonic map flow into spheres. *Math. Comp.*, 76(260):1847–1859 (electronic), 2007.

[13] P. Bénilan, L. Boccardo, T. Gallouët, R. Gariepy, M. Pierre, and J. L. Vázquez. An L^1-theory of existence and uniqueness of solutions of nonlinear elliptic equations. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 22(2):241–273, 1995.

[14] M. Bessemoulin-Chatard, C. Chainais-Hillairet, and F. Filbet. On discrete functional inequalities for some finite volume schemes. *IMA Journal of Numerical Analysis*, 2014.

[15] A. Bressan and P. LeFloch. Uniqueness of weak solutions to systems of conservation laws. *Arch. Rational Mech. Anal.*, 140(4):301–317, 1997.

[16] A. Bressan, T.-P. Liu, and T. Yang. L^1 stability estimates for $n \times n$ conservation laws. *Arch. Ration. Mech. Anal.*, 149(1):1–22, 1999.

[17] H. Brinkman. A calculation of the viscous force exerted by a flowing fluid on a dense swarm of particles. *Applied Scientific Research*, 1(1):27–34, 1949.

[18] W. Brown. *Micromagnetics*. Interscience tracts on physics and astronomy. Interscience Publishers, 1963.

[19] A. Brú, S. Albertos, J. L. Subiza, J. L. García-Asenjo, and I. Brú. The universal dynamics of tumor growth. *Biophysical journal*, 85(5):2948–2961, 2003.

[20] H. Byrne and D. Drasdo. Individual-based and continuum models of growing cell populations: a comparison. *J. Math. Biol.*, 58(4-5):657–687, 2009.

[21] H. Byrne and L. Preziosi. Modelling solid tumour growth using the theory of mixtures. *Mathematical Medicine and Biology*, 20(4):341–366, 2003.

[22] M. C. Calderer, D. Golovaty, F. H. Lin, and C. Liu. Time evolution of nematic liquid crystals with variable degree of orientation. *SIAM Journal on Mathematical Analysis*, 33(5):1033–1047, 2002.

[23] J. Casado-Díaz, T. Chacón Rebollo, V. Girault, M. Gómez Marmol, and F. Murat. Finite elements approximation of second order linear elliptic equations in divergence form with right-hand side in L^1. *Numerische Mathematik*, 105(3):337–374, 2007.

[24] M. A. J. Chaplain. Avascular growth, angiogenesis and vascular growth in solid tumours: The mathematical modelling of the stages of tumour development. *Mathematical and computer modelling*, 23(6):47–87, 1996.

[25] N. Chemetov and W. Neves. The generalized Buckley-Leverett system: solvability. *Arch. Ration. Mech. Anal.*, 208(1):1–24, 2013.

[26] D. Chen and A. Friedman. A two-phase free boundary problem with discontinuous velocity: application to tumor model. *J. Math. Anal. Appl.*, 399(1):378–393, 2013.

[27] G. M. Coclite, L. di Ruvo, J. Ernest, and S. Mishra. Convergence of vanishing capillarity approximations for scalar conservation laws with discontinuous fluxes. *Netw. Heterog. Media*, 8(4):969–984, 2013.

[28] G. M. Coclite, H. Holden, and K. H. Karlsen. Wellposedness for a parabolic-elliptic system. *Discrete Contin. Dyn. Syst.*, 13(3):659–682, 2005.

[29] G. M. Coclite, K. H. Karlsen, S. Mishra, and N. H. Risebro. A hyperbolic-elliptic model of two-phase flow in porous media—existence of entropy solutions. *Int. J. Numer. Anal. Model.*, 9(3):562–583, 2012.

[30] G. M. Coclite, S. Mishra, and N. H. Risebro. Convergence of an Engquist-Osher scheme for a multidimensional triangular system of conservation laws. *Math. Comp.*, 79(269):71–94, 2010.

[31] G. M. Coclite, S. Mishra, N. H. Risebro, and F. Weber. Analysis and numerical approximation of Brinkman regularization of two-phase flows in porous media. *Comput. Geosci.*, 18(5):637–659, 2014.

[32] G. M. Coclite and N. H. Risebro. Conservation laws with time dependent discontinuous coefficients. *SIAM J. Math. Anal.*, 36(4):1293–1309 (electronic), 2005.

[33] S. Cui. Formation of necrotic cores in the growth of tumors: Analytical results. *Acta Mathematica Scientia*, 26(4):781 – 796, 2006.

[34] S. Cui and J. Escher. Asymptotic behaviour of solutions of a multidimensional moving boundary problem modeling tumor growth. *Communications in Partial Differential Equations*, 33(4):636–655, 2008.

[35] S. Cui and A. Friedman. Analysis of a mathematical model of the growth of necrotic tumors. *Journal of Mathematical Analysis and Applications*, 255(2):636 – 677, 2001.

[36] C. M. Dafermos. Polygonal approximations of solutions of the initial value problem for a conservation law. *J. Math. Anal. Appl.*, 38:33–41, 1972.

[37] C. M. Dafermos. *Hyperbolic conservation laws in continuum physics*, volume 325 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 2010.

[38] R. J. DiPerna and P.-L. Lions. Ordinary differential equations, transport theory and Sobolev spaces. *Invent. Math.*, 98(3):511–547, 1989.

[39] D. Donatelli and K. Trivisa. On a nonlinear model for the evolution of tumor growth with a variable total density of cancerous cells. Submitted, 2014.

[40] D. Donatelli and K. Trivisa. On a nonlinear model for the evolution of tumor growth with drug application, 2014. Submittted to Nonlinearity.

[41] D. Donatelli and K. Trivisa. On a nonlinear model for tumor growth: global in time weak solutions. *J. Math. Fluid Mech.*, 16(4):787–803, 2014.

[42] D. Drasdo and S. Höhme. A single-cell-based model of tumor growth in vitro : monolayers and spheroids. *Physical Biology*, 2(3):133, 2005.

[43] M. Dreher and A. Jüngel. Compact families of piecewise constant functions in $L^p(0,T;B)$. *Nonlinear Anal.*, 75(6):3072–3077, 2012.

[44] W. E, K. Khanin, A. Mazel, and Y. Sinai. Invariant measures for Burgers equation with stochastic forcing. *Ann. of Math. (2)*, 151(3):877–960, 2000.

[45] T. Elperin, N. Kleeorin, and A. Krylov. Nondissipative shock waves in two-phase flows. *Physica D*, 74(3-4):372–385, July 1994.

[46] L. C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010.

[47] L. C. Evans and R. F. Gariepy. *Measure theory and fine properties of functions*. Studies in advanced mathematics. CRC Press, Boca Raton (Fla.), 1992.

[48] R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In *Handbook of numerical analysis, Vol. VII*, Handb. Numer. Anal., VII, pages 713–1020. North-Holland, Amsterdam, 2000.

[49] R. Eymard, T. Gallouët, R. Herbin, and J. C. Latché. A convergent finite element-finite volume scheme for the compressible Stokes problem. II. The isentropic case. *Math. Comp.*, 79(270):649–675, 2010.

[50] F. C. Frank. I. Liquid crystals. On the theory of liquid crystals. *Discuss. Faraday Soc.*, 25:19–28, 1958.

[51] A. Friedman. A hierarchy of cancer models and their mathematical challenges. *Discrete Contin. Dyn. Syst. Ser. B*, 4(1):147–159, 2004. Mathematical models in cancer (Nashville, TN, 2002).

[52] A. Friedman and B. Hu. Stability and instability of liapunov-schmidt and hopf bifurcation for a free boundary problem arising in a tumor model. *Transactions of the American Mathematical Society*, 360(10):5291–5342, 2008.

[53] D. Gilbarg and N. S. Trudinger. *Elliptic partial differential equations of second order*. Classics in Mathematics. Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

[54] M. B. Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In *Monte Carlo and quasi-Monte Carlo methods 2006*, pages 343–358. Springer, Berlin, 2008.

[55] M. B. Giles. Multilevel Monte Carlo path simulation. *Oper. Res.*, 56(3):607–617, 2008.

[56] T. Gimse and N. H. Risebro. Solution of the Cauchy Problem for a Conservation Law with a Discontinuous Flux Function. *SIAM Journal on Mathematical Analysis*, 23(3):635–648, 1992.

[57] E. Giusti. *Minimal surfaces and functions of bounded variation*, volume 80 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 1984.

[58] E. Godlewski and P.-A. Raviart. *Hyperbolic systems of conservation laws*, volume 3/4 of *Mathématiques & Applications (Paris) [Mathematics and Applications]*. Ellipses, Paris, 1991.

[59] E. Godlewski and P.-A. Raviart. *Numerical approximation of hyperbolic systems of conservation laws*, volume 118 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996.

[60] R. L. Graham and F. F. Yao. Finding the convex hull of a simple polygon. *J. Algorithms*, 4(4):324–331, 1983.

[61] H. P. Greenspan. Models for the growth of a solid tumor by diffusion. *Stud. Appl. Math*, 51(4):317–340, 1972.

[62] H. P. Greenspan. On the growth and stability of cell cultures and solid tumors. *J. Theoret. Biol.*, 56(1):229–242, 1976.

[63] S. M. Hassanizadeh and W. G. Gray. Mechanics and thermodynamics of multiphase flow in porous media including interphase boundaries. *Advances in water resources*, 13(4):169–186, 1990.

[64] S. Heinrich. Multilevel Monte Carlo methods. In *Large–scale scientific computing*, pages 58–67. Springer, Berlin, 2001.

[65] F. Hélein. *Harmonic maps, conservation laws and moving frames*, volume 150 of *Cambridge Tracts in Mathematics*. Cambridge University Press, Cambridge, second edition, 2002. Translated from the 1996 French original, With a foreword by James Eells.

[66] F. Hélein and J. C. Wood. Harmonic maps. In *Handbook of global analysis*, pages 417–491, 1213. Elsevier Sci. B. V., Amsterdam, 2008.

[67] R. Helmig, A. Weiss, and B. I. Wohlmuth. Dynamic capillary effects in heterogeneous porous media. *Comput. Geosci.*, 11(3):261–274, 2007.

[68] H. Holden and L. Holden. On scalar conservation laws in one dimension. In *Ideas and methods in mathematical analysis, stochastics, and applications (Oslo, 1988)*, pages 480–509. Cambridge Univ. Press, Cambridge, 1992.

[69] H. Holden, L. Holden, and R. Høegh-Krohn. A numerical method for first order nonlinear scalar conservation laws in one dimension. *Comput. Math. Appl.*, 15(6-8):595–602, 1988. Hyperbolic partial differential equations. V.

[70] H. Holden, K. H. Karlsen, K.-A. Lie, and N. H. Risebro. *Splitting methods for partial differential equations with rough solutions*. EMS Series of Lectures in Mathematics. European Mathematical Society (EMS), Zürich, 2010. Analysis and MATLAB programs.

[71] H. Holden, T. Lindstrøm, B. Øksendal, J. Ubøe, and T.-S. Zhang. The Burgers equation with a noisy force and the stochastic heat equation. *Comm. Partial Differential Equations*, 19(1-2):119–141, 1994.

[72] H. Holden and N. H. Risebro. Conservation laws with a random source. *Appl. Math. Optim.*, 36(2):229–241, 1997.

[73] H. Holden and N. H. Risebro. *Front tracking for hyperbolic conservation laws*, volume 152 of *Applied Mathematical Sciences*. Springer, New York, 2011. First softcover corrected printing of the 2002 original.

[74] K. H. Karlsen and T. K. Karper. A convergent nonconforming finite element method for compressible Stokes flow. *SIAM J. Numer. Anal.*, 48(5):1846–1876, 2010.

[75] K. H. Karlsen and T. K. Karper. Convergence of a mixed method for a semi-stationary compressible Stokes system. *Math. Comp.*, 80(275):1459–1498, 2011.

[76] K. H. Karlsen and T. K. Karper. A convergent mixed method for the Stokes approximation of viscous compressible flow. *IMA J. Numer. Anal.*, 32(3):725–764, 2012.

[77] K. H. Karlsen, N. H. Risebro, and J. D. Towers. L^1 stability for entropy solutions of nonlinear degenerate parabolic convection-diffusion equations with discontinuous coefficients. *Skr. K. Nor. Vidensk. Selsk.*, (3):1–49, 2003.

[78] T. K. Karper. A convergent FEM-DG method for the compressible Navier-Stokes equations. *Numer. Math.*, 125(3):441–510, 2013.

[79] T. K. Karper and F. Weber. A new angular momentum method for computing wave maps into spheres. *SIAM J. Numer. Anal.*, 52(4):2073–2091, 2014.

[80] F. Kissling, R. Helmig, and C. Rohde. Simulation of infiltration processes in the unsaturated zone using a multi-scale approach. *Vadose Zone J.*, 11(3):–, 2012.

[81] F. Kissling and K. H. Karlsen. On the singular limit of a two-phase flow equation with heterogeneities and dynamic capillary pressure. *ZAMM Z. Angew. Math. Mech.*, 94(7-8):678–689, 2014.

[82] J. Krieger and W. Schlag. *Concentration compactness for critical wave maps*. EMS Monographs in Mathematics. European Mathematical Society (EMS), Zürich, 2012.

[83] P. S. Krishnaprasad and X. Tan. Cayley transforms in micromagnetics. *Physica B: Condensed Matter*, 306(1):195–199, 2001.

[84] M. Krotkiewski, I. S. Ligaarden, K.-A. Lie, and D. W. Schmid. On the Importance of the Stokes-Brinkman Equations for Computing Effective Permeability in Karst Reservoirs. *Communications in Computational Physics*, 10(5):1315–1332, 2011.

[85] S. N. Kružkov. First order quasilinear equations with several independent variables. *Mat. Sb. (N.S.)*, 81 (123):228–255, 1970.

[86] S. N. Kružkov and S. M. Sukorjanskiĭ. Boundary value problems for systems of equations of two-phase filtration type; formulation of problems, questions of solvability, justification of approximate methods. *Mat. Sb. (N.S.)*, 104(146)(1):69–88, 175–176, 1977.

[87] O. A. Ladyzhenskaya. *The mathematical theory of viscous incompressible flow*. Second English edition, revised and enlarged. Translated from the Russian by Richard A. Silverman and John Chu.

Mathematics and its Applications, Vol. 2. Gordon and Breach, Science Publishers, New York-London-Paris, 1969.

[88] O. A. Ladyzhenskaya. *The boundary value problems of mathematical physics*, volume 49 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1985. Translated from the Russian by Jack Lohwater [Arthur J. Lohwater].

[89] O. A. Ladyzhenskaya, V. A. Solonnikov, and N. N. Ural′ceva. *Linear and quasilinear equations of parabolic type*. Translated from the Russian by S. Smith. Translations of Mathematical Monographs, Vol. 23. American Mathematical Society, Providence, R.I., 1968.

[90] P. G. LeFloch. *Hyperbolic systems of conservation laws*. Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, 2002. The theory of classical and nonclassical shock waves.

[91] R. J. LeVeque. *Finite volume methods for hyperbolic problems*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2002.

[92] P.-L. Lions. *Mathematical topics in fluid mechanics. Vol. 1*, volume 3 of *Oxford Lecture Series in Mathematics and its Applications*. The Clarendon Press, Oxford University Press, New York, 1996. Incompressible models, Oxford Science Publications.

[93] P.-L. Lions. *Mathematical topics in fluid mechanics. Vol. 2*, volume 10 of *Oxford Lecture Series in Mathematics and its Applications*. The Clarendon Press, Oxford University Press, New York, 1998. Compressible models, Oxford Science Publications.

[94] S. Lukkhaus and P. I. Plotnikov. Entropy solutions of Buckley-Leverett equations. *Sibirsk. Mat. Zh.*, 41(2):400–420, iv, 2000.

[95] E. Marušić-Paloka, I. Pažanin, and S. Marušić. Comparison between Darcy and Brinkman laws in a fracture. *Appl. Math. Comput.*, 218(14):7538–7545, 2012.

[96] S. Mishra, N. H. Risebro, C. Schwab, and S. Tokareva. Numerical solution of scalar conservation laws with random flux functions. Technical Report 2012-35, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2012.

[97] S. Mishra and C. Schwab. Sparse tensor multi-level Monte Carlo finite volume methods for hyperbolic conservation laws with random initial data. *Mathematics of Computation*, 81(280):1979–2018, 2012.

[98] S. Mishra, C. Schwab, and J. Šukys. Multi-level Monte Carlo finite volume methods for nonlinear systems of conservation laws in multi-dimensions. *Journal of Computational Physics*, 231(8):3365–3388, 2012.

[99] S. Mishra, C. Schwab, and J. Šukys. Multi-level Monte Carlo finite volume methods for uncertainty quantification in nonlinear systems of balance laws. In *Uncertainty Quantification in Computational Fluid Dynamics*, pages 225–294. Springer, 2013.

[100] S. P. Neuman. Theoretical derivation of Darcy's law. *Acta Mechanica*, 25(3-4):153–170, 1977.

[101] A. Novotný and I. Straškraba. *Introduction to the mathematical theory of compressible flow*, volume 27 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2004.

[102] F. Otto. *Stability Investigation of Planar Solutions of the Buckley-Leverett Equations*. Preprint: Sonderforschungsbereich Nichtlineare Partielle Differentialgleichungen. Sonderforschungsbereich 256, 1994.

[103] B. Perthame. *Kinetic formulation of conservation laws*. Oxford lecture series in mathematics and its applications. Oxford university press, New York, 2002.

[104] B. Perthame, F. Quirós, M. Tang, and N. Vauchelet. Derivation of a Hele–Shaw type system from a cell model with active motion. *Interfaces Free Bound.*, 16(4):489–508, 2014.

[105] B. Perthame, F. Quirós, and J. L. Vázquez. The Hele-Shaw asymptotics for mechanical models of tumor growth. *Arch. Ration. Mech. Anal.*, 212(1):93–127, 2014.

[106] B. Perthame, M. Tang, and N. Vauchelet. Traveling wave solution of the Hele-Shaw model of tumor growth with nutrient. *Math. Models Methods Appl. Sci.*, 24(13):2601–2626, 2014.

[107] B. Perthame and N. Vauchelet. Incompressible limit of mechanical model of tumor growth with viscosity. Preprint, 2014.

[108] J. Ranft, M. Basan, J. Elgeti, J.-F. Joanny, J. Prost, and F. Jülicher. Fluidization of tissues by cell division and apoptosis. *Proceedings of the National Academy of Sciences*, 107(49):20863–20868, 2010.

[109] N. H. Risebro, C. Schwab, and F. Weber. Multilevel Monte Carlo front-tracking for random scalar conservation laws. *BIT Numerical Mathematics*, pages 1–30, 2015.

[110] R. A. Saxton. Dynamic instability of the liquid crystal director. In *Current progress in hyperbolic systems: Riemann problems and computations (Brunswick, ME, 1988)*, volume 100 of *Contemp. Math.*, pages 325–330. Amer. Math. Soc., Providence, RI, 1989.

[111] J. Shatah and M. Struwe. *Geometric wave equations*, volume 2 of *Courant Lecture Notes in Mathematics*. New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 1998.

[112] D. Speiser, K. Williams, and S. Caparrini. *Discovering the Principles of Mechanics 1600-1800: Essays by David Speiser*. Essays by David Speiser. Springer Basel AG, 2008.

[113] I. W. Stewart. *The Static and Dynamic Continuum Theory of Liquid Crystals: A Mathematical Introduction*. CRC Press, 2004.

[114] T. Tao. Ill-posedness for one-dimensional wave maps at the critical regularity. *Amer. J. Math.*, 122(3):451–463, 2000.

[115] T. Tao. Global regularity of wave maps. I. Small critical Sobolev norm in high dimension. *Internat. Math. Res. Notices*, (6):299–328, 2001.

[116] T. Tao. Global regularity of wave maps. II. Small energy in two dimensions. *Comm. Math. Phys.*, 224(2):443–544, 2001.

[117] T. Tao. *Nonlinear dispersive equations*, volume 106 of *CBMS Regional Conference Series in Mathematics*. Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 2006. Local and global analysis.

[118] D. Tataru. The wave maps equation. *Bull. Amer. Math. Soc. (N.S.)*, 41(2):185–204 (electronic), 2004.

[119] D. Tataru. Rough solutions for the wave maps equation. *Amer. J. Math.*, 127(2):293–377, 2005.

[120] C. J. van Duijn, Y. Fan, L. A. Peletier, and I. S. Pop. Travelling wave solutions for degenerate pseudo-parabolic equations modelling two-phase flow in porous media. *Nonlinear Anal. Real World Appl.*, 14(3):1361–1383, 2013.

[121] C. J. van Duijn, L. A. Peletier, and I. S. Pop. A new class of entropy solutions of the Buckley-Leverett equation. *SIAM J. Math. Anal.*, 39(2):507–536 (electronic), 2007.

[122] J. van Neerven. Stochastic evolution equations. Lecture Notes, ISEM, 2007/8.

[123] J. L. Vázquez. *The porous medium equation*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, Oxford, 2007. Mathematical theory.

[124] J. von Neumann. The Mathematician. In R. B. Heywood, editor, *The Works of the Mind*. University of Chicago Press, Chicago, 1947.

[125] J. Wehr and J. Xin. White noise perturbation of the viscous shock fronts of the Burgers equation. *Comm. Math. Phys.*, 181(1):183–203, 1996.

[126] J. Wehr and J. Xin. Front speed in the Burgers equation with a random flux. *J. Statist. Phys.*, 88(3-4):843–871, 1997.

[127] K. Widmayer. Non-uniqueness of weak solutions to the wave map problem. *Annales de l'Institut Henri Poincare (C) Non Linear Analysis*, (0):–, 2014.

[128] J.-H. Zhao. A parabolic-hyperbolic free boundary problem modeling tumor growth with drug application. *Electron. J. Differential Equations*, pages No. 03, 18, 2010.